# INTRODUCTION TO
# ECONOMIC STATISTICS

# INTRODUCTION TO
# ECONOMIC STATISTICS

BY

GEORGE R. DAVIES, Ph. D.

PROFESSOR OF SOCIOLOGY, UNIVERSITY OF NORTH DAKOTA

NEW YORK
THE CENTURY CO.
1922

# PREFACE

The application of statistical methods to special fields such as demography, education, and economics has in recent years advanced very rapidly. As a result a study of statistical principles must be confined chiefly to one of the special fields if it is not to be lost in the multiplicity of specific methods and illustrations. This text-book has been written with the interests of the student of economics in mind.

A common difficulty which the teacher of statistics encounters is a lack of provision for laboratory work. An attempt has here been made to supply this need by furnishing illustrative problems, graphs, and data which may be worked over by the student, and by adding to each chapter a list of related exercises. The exercises will be found extensive enough so that the teacher may select those which are adapted to his requirements. The longer problems should be subdivided, and the parts assigned to different members of the class. Both the tables and the exercises may very well be supplemented by the use of data drawn from such sources as the *Survey of Current Business,* the *Monthly Labor Review,* and the *Statistical Abstract of the United States* (Superintendent of Documents, Government Printing Office, Washington, D. C.). The topics covered in the text represent probably a maximum of what can be mastered by a college class in a term. Perhaps it may be found advisable to omit certain topics, such as interpolating for quartiles, theories

of price indexes, parabola trends, seasonal variations, and the more complex methods of correlation.

Nearly all the material here presented has been accumulated from experience in the statistical laboratory and class-room. Particular attention has been given to the requirements in respect to fundamental theory of the statistical departments of the larger banks and business houses. Some of the recently developed methods of handling business barometers have therefore been touched upon, and some attention has been given to the theory of price and production indexes.

The book is an outgrowth of the undergraduate course in statistics given by the writer at Princeton University during the school year 1920-1921. This course was modeled in its general features upon the course given the preceding year by Professor J. H. Williams, now of Harvard University. The writer wishes to acknowledge his indebtedness to Professor Williams for the general plan of the laboratory exercises, as well as for many specific suggestions. Thanks are also extended to Professors F. A. Fetter and E. W. Kemmerer of Princeton for their interest and encouragement, and to Professor W. F. Willcox of Cornell who read the first draft of the manuscript and made several valuable suggestions for its revision. Indebtedness is also acknowledged to the following for their kind permission to reprint data: Mr. Roger W. Babson, Professor Stanley E. Howard, Bureau of Labor Statistics, National City Bank, National Bureau of Economic Research, National Industrial Conference Board, *Review of Economic Statistics,* and the *Quarterly Journal* of the University of North Dakota.

# CONTENTS

INTRODUCTION TO
ECONOMIC STATISTICS

# INTRODUCTION TO ECONOMIC STATISTICS

## CHAPTER I

## TABULATION

The term "statistics" when used to designate a branch of study, implies an exposition of certain methods employed in presenting and interpreting the numerical aspects of a given subject. The science of statistics consists, therefore, of principles and methods, rather than of data. The principles are essentially the same whether the application is made to biology, demography, education, or economics. But the detailed methods in these and other fields have in late years become so specialized that it is hardly practicable any longer to study statistics in the abstract. The field of application here adopted is chiefly that of general economics.[1] Illustrations will be given of the methods employed in organizing data, in computing and employing indexes, and in measuring trends and correlations.

[1] It should be noted that economic statistics are commonly distinguished from business statistics. The former subject studies general market conditions, while the latter subject deals with the details of a specific business establishment, and is therefore an adjunct of accounting. Business statistics vary so greatly from one establishment to another that it is difficult to generalize from them. Their problems consist largely of the application of statistical principles to specific situations. For a discussion of the distinction see "The Scope of Business Statistics," by R. P. Falkner, in the *Quarterly Publications of the American Statistical Association*, June, 1918, pp. 24-29.

**Preliminary Schedules.**  Statistical field work commonly begins with the preparation of schedules in which to enter the desired data.  These schedules may be in the nature of questionnaires, or they may be mere forms in which to copy certain records.  Since such schedules will vary with the particular task in hand, few rules can be laid down for their preparation.  Experience will teach, however, that they must be very carefully worked out in advance.  In the first place, it must be determined as precisely as possible what data will be needed.  If the schedule is in questionnaire form, great care must be taken to make the questions unambiguous and easily comprehended by those who are to answer.  Errors may often be checked by asking a question in two ways, at different places in the list; as by calling for both the age and the date of birth.

After the preliminary schedules have been compiled, the statistician begins to organize his data, so that conclusions may be deduced and presented in simplified form.  In so doing he will make use of the processes of tabulation.  These may best be explained by taking an example.  As such an example, we shall choose certain wage schedules which have been recorded in the Aldrich Report on ''Wholesale Prices, Wages, and Transportation'' (Senate Report No. 1394, dated 1893).  The data here used will be found in Vol. IV, pages 1463-1497, and refer to a Connecticut woolen mill designated as Establishment No. 86.  The wage rolls as recorded in the report cover about half a century, ending in 1891.  They show the daily wages paid to each class of workmen employed in the mill in January and July of each year.  We shall select

for tabulation only the wages paid in July of 1870, 1880, and 1891.

**The Primary or General Purpose Table.** In transcribing the selected wage schedules into a primary table, it will be found advisable to edit the original figures by making certain minor modifications. An inspection of the schedules will show that most of the wage rates are expressed as multiples of five cents. Of course, we may assume theoretically that the exact economic values which are approximated in the actual wages must form a continuous series, instead of one having a regular interval. That is, if the wages could be expressed as theoretically exact values, and if a very large number of workers were involved, the rates would be separated by intervals of only a small fraction of a cent. The case may be compared with the measurement of the height of the individuals in a large group. If the measurements are taken with very precise instruments, the results will be expressed in hundredths or perhaps thousandths of an inch. But for practical purposes measurements to perhaps the nearest quarter inch are sufficiently accurate. In the same way when the employer made an offer of wages he set his figure at a multiple of five cents, or in the case of the larger wages at a multiple of twenty-five cents. For his purpose, such an estimate of the market value of labor was sufficiently accurate. The exceptions will be found to be rates that were paid to an especially large number of workers. In such cases differences of a cent or two are of considerable consequence.

**Continuous and Discrete Series.** A series of measurements or values occurring only at more or less reg-

ular intervals is said to be discrete. Sometimes a series
will be naturally discrete, as when flowers are clas-
sified by the number of petals, or when the spots in
a number of dice-throws are tabulated. But a series
which is theoretically continuous becomes artificially
discrete when a limit of accuracy is determined upon,
as when height is measured to the nearest quarter-
inch, or wages are expressed in multiples of five cents.
Our series of wage rates, as it stands in the records,
is therefore discrete at intervals of five cents, except
for a few items.

For purposes of classification it is desirable to have
our series of wage rates regularly discrete throughout.
Since it is not possible to break down the five cent
intervals to smaller ones, it will be necessary to modify
such rates as 67c. and $1.28 so as to classify them as
multiples of five. If there were only a few scattered
cases of this sort affecting only a small number of
workers, they might merely be entered at the nearest
five cent intervals; that is, 67c. could be entered as
65c. and $1.28 as $1.30. But since the number of work-
ers at these irregular rates is exceptionally large,
such a procedure might result in too great a degree of
inaccuracy. We shall therefore apply a familiar arith-
metical device and break up the given number of work-
ers at each inconvenient rate into two groups, one
having a higher and one a lower rate than that stated,
but maintaining together the same average wage. The
procedure may be illustrated as follows:

$$15 \text{ workers at } 67c. = \begin{cases} 9 \text{ workers at } 65c. \\ + \\ 6 \text{ workers at } 70c. \end{cases}$$

This result is obtained by taking 3/5 and 2/5 of the fifteen workers (the fractional parts of the five cent interval which lie between 67c. and the nearest multiples of 5) and placing the larger of the two results at the rate expressed by the multiple of five nearest to 67c., that is, at 65c. The smaller of the two results is

TABLE I

WAGE ROLL IN A CONNECTICUT WOOLEN MILL
JULY OF SPECIFIED YEARS

| OCCUPATION | 1870 | | 1880 | | 1891 | |
|---|---|---|---|---|---|---|
| | NO. | WAGE | NO. | WAGE | NO. | WAGE |
| Burlers...................... | 1* | $ .80 | 11* | $ .80 | 14* | $1.10 |
| | 2* | .85 | 8* | .85 | | |
| Card cleaners................ | 1 | .70 | 4 | 1.15 | 1 | 1.10 |
| | 2 | 1.10 | | | 1 | 1.15 |
| | 1 | 1.25 | | | 1 | 1.20 |
| | | | | | 2 | 1.25 |
| Card tenders................ | 1 | .55 | 2 | .60 | 1 | .75 |
| | 2 | .60 | 2 | .65 | 1 | .85 |
| | 3 | .65 | | | 1 | .90 |
| | 1 | .80 | | | 1 | 1.00 |
| Carpenters.................. | 1 | 2.75 | 1 | 2.75 | 1 | 2.75 |
| Cloth inspectors.............. | | | 1 | 1.60 | 1 | 1.90 |
| Drawers-in.................. | 1* | 1.25 | 1* | 1.90 | 1* | 1.75 |
| | | | 1* | 1.95 | 1* | 1.85 |
| | | | | | 4* | 1.90 |
| Dressers..................... | | | 2 | 1.50 | 1 | 1.25 |
| | | | 2 | 1.55 | 2 | 1.60 |
| | | | | | 1 | 1.65 |
| | | | | | 3 | 1.75 |
| Dyers....................... | 2 | 1.35 | 3 | 1.25 | 1 | 1.15 |
| | 1 | 1.50 | 5 | 1.30 | 13 | 1.25 |
| | | | | | 1 | 1.40 |
| | | | | | 2 | 1.50 |
| Firemen..................... | 1 | 1.50 | | | 3 | 1.50 |
| Foremen-burlers.............. | | | | | 1 | 2.25 |
| Fullers and giggers........... | 4 | 1.10 | 2 | 1.05 | 2 | 1.05 |
| | 3 | 1.15 | 4 | 1.20 | 4 | 1.10 |
| | 2 | 1.25 | 1 | 1.25 | 3 | 1.15 |
| | 1 | 1.35 | | | 3 | 1.25 |
| | 2 | 1.50 | | | 1 | 1.50 |
| Handers-in.................. | 2* | .40 | 2* | .40 | 5* | .50 |
| Harness hands.............. | | | | | 1 | 1.50 |
| Loom fixers................. | 2 | 1.50 | 3 | 1.95 | 1 | 2.00 |
| | 1 | 1.10 | | | 2 | 2.10 |
| | 1 | 1.15 | | | 4 | 2.20 |

TABLE I (Continued)

| OCCUPATION | 1870 | | 1880 | | 1891 | |
|---|---|---|---|---|---|---|
| | NO. | WAGE | NO. | WAGE | NO. | WAGE |
| Machinists................... | 1 | $2.75 | 1 | $3.00 | 1 | $2.75 |
| Machinists—helpers........... | 1 | 1.75 | | | 1 | 2.10 |
| Number sewers.............. | | | | | 3* | .90 |
| Overseers—Carding dept...... | 1 | 3.50 | 1 | 3.25 | 1 | 4.00 |
| "        Dyehouse......... | 1 | 3.75 | 1 | 3.25 | 1 | 4.25 |
| "        Finishing dept..... | 1 | 2.50 | 1 | 3.00 | 1 | 3.50 |
| "        Fulling dept....... | 1 | 2.75 | 1 | 2.50 | 1 | 2.50 |
| "        Spinning dept..... | 1 | 2.75 | 1 | 3.00 | 1 | 2.75 |
| "        Spooling dept...... | 1 | 2.25 | 1 | 2.25 | 1 | 2.50 |
| "        Weaving dept...... | 1 | 3.00 | 1 | 3.00 | 3 | 3.00 |
| Piecers..................... | 1 | .75 | 2 | .70 | | |
| | 1 | .80 | 8 | .75 | | |
| Second hands................ | 1 | 1.50 | 1 | 1.25 | 2 | 1.75 |
| | 1 | 1.75 | 1 | 1.50 | | |
| Sewers..................... | | | 1* | 1.25 | 1* | 1.00 |
| | | | | | 2* | 1.25 |
| Shearers.................... | 1 | 1.15 | 6 | 1.25 | 8 | 1.35 |
| | 1 | 1.40 | | | | |
| | 1 | 1.50 | | | | |
| Sorters..................... | 2 | 2.00 | 2 | 1.80 | 2 | 1.70 |
| | 1 | 2.75 | 1 | 2.75 | 1 | 2.75 |
| Speckers.................... | 1* | 1.35 | 1* | .90 | 2* | .80 |
| | | | 2* | .95 | 2* | .85 |
| Spinners—jack and mule...... | 3 | 1.75 | 1 | 1.50 | 4 | 1.25 |
| | 7 | 1.80 | | | 7 | 1.30 |
| Spoolers.................... | 5* | .60 | 5* | .65 | 9* | .75 |
| | | | 4* | .70 | 14* | .80 |
| Teamsters................... | 1 | 1.50 | 1 | 1.50 | 1 | 1.60 |
| Twisters.................... | | | 7* | .90 | 4* | .90 |
| Watchmen................... | 1 | 1.50 | 1 | 1.50 | 1 | 1.30 |
| | | | | | 1 | 1.35 |
| | | | | | 1 | 1.40 |
| Weavers.................... | 4* | 1.05 | 79 | 1.20 | 2* | 1.30 |
| | 3* | 1.10 | 10* | 1.40 | 8* | 1.35 |
| | 19 | 1.30 | 2* | 1.45 | 89 | 1.70 |
| | 5 | 1.35 | | | 60 | 1.75 |
| Weavers—pattern............ | | | | | 3 | 1.25 |
| | | | | | 2 | 1.35 |
| | | | | | 4 | 1.50 |
| | | | | | 1 | 1.75 |
| | | | | | 1 | 2.00 |
| Winders.................... | 1* | 1.00 | 6* | .95 | 12* | 1.15 |
| | | | 8* | 1.00 | 17* | 1.20 |
| Yarn carriers............... | 1 | 1.25 | 1 | 1.75 | 1 | 1.85 |
| Totals................... | 108 | | 213 | | 361 | |

* Female employes.

placed at the rate of 70c. That the average wage has not been changed by this operation is shown by the fact that

$$15 \times 67c. = 9 \times 65c + 6 \times 70c.$$

By the foregoing method all irregular wage rates may now be reduced to approximately equivalent multiples of five cents. Of course, when a fraction appears in the operation the nearest whole number is taken. Thus modified, our wage schedules appear as shown in Table I, which may be taken as an example of a primary or general purpose table.

**The Frequency Curve.** The tabulation which we are about to undertake has for its immediate object the presentation of the frequency distribution of the wages in question. Since the concept of a frequency distribution has a concise theoretical basis, it will be of advantage to turn briefly at this point to the theoretical aspects of the subject.

If we take the square of a binomial, as $a^2 + 2ab + b^2$, we have three classes of values as expressed by the letters and their exponents, and these classes have frequencies expressed by their coefficients, $1:2:1$. If instead of the second power we take the fourth power of the binomial, we have five classes of values, having frequencies respectively of $1:4:6:4:1$. These frequencies graphed as vertical blocks will form a figure such as is outlined by the dotted line in Fig. 1. If instead of the fourth power of the binomial we should take the thousandth or millionth power, the steps in this blocked frequency polygon would practically disappear, and the figure would approach a smoothed bell-shaped curve as indicated in the same figure. This theoretical

distribution of classes of values, or something similar to it, may be discovered to exist very generally in natural and social phenomena, and is also the expression of what are known as the laws of chance. The length of leaves on a given tree, the height of a group of persons, the per cent net earnings of corporations,
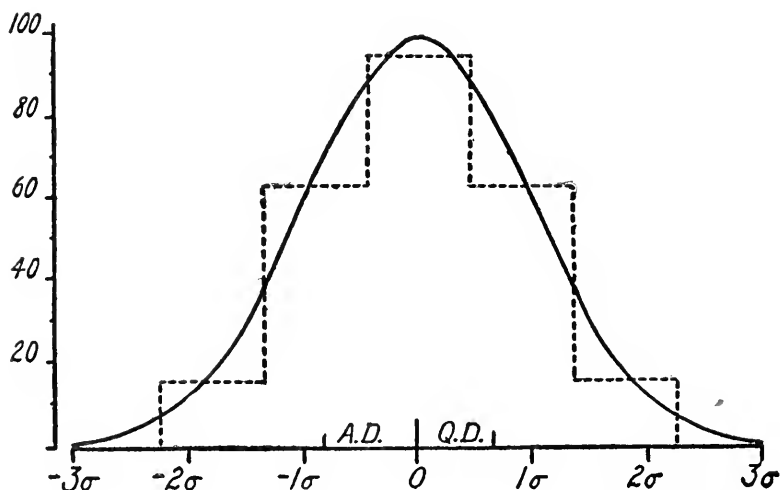


Figure 1. Normal frequency curve (solid line), and an approximation to it (dotted line) based on the fourth power of a bionomial. Horizontal scale in units of standard deviation from average (0). Quartile deviation (Q.D.), and average deviation (A.D.)

or the deviations from normal of a price index through a series of years, will show when properly classified and graphed an approximation to the bell-shaped frequency curve. In order to discover whether this curve is inherent in a given set of data, it is necessary first that the data contain a considerable number of items, and second that the classification be suitably adjusted to the range and numbers. If, for example, the height of a hundred persons were taken merely to the nearest

foot, only two or three classes would appear. If, how-
ever, the measurements were taken accurately to .01
inch, so many classes would appear that the fre-
quencies would be hopelessly scattered. But if we
made our measurements to the nearest inch, we would
obtain a series of frequencies somewhat like the fol-
lowing (the classes ranging from 60 to 73 inches in-
clusive) : 1 :2 :4 :7 :10 :14 :16 :16 :12 :8 :5 :3 :1 :1. These fre-
quencies when graphed will give an approximation to
the bell-shaped curve. It will be necessary, therefore,
in tabulating our wage data to work out experimentally
the most suitable classification.

**The Tally Sheet and Frequency Table.** To facilitate
the classification of the wages selected for study, a tally
sheet is drawn up as shown in the first two columns of
Table II. We shall show here the details for only the
1891 figures, leaving the 1870 and 1880 figures to be
worked out by the•student. In studies where the items
must be entered singly, it is customary to use the famil-
iar ''four and cross'' method of tallying ($\mathcal{HH} = 5$),
but this is inappropriate when the items are already
partially grouped as they are here. In this case the
number of workers as shown in the wage roll is entered
in the appropriate line of the tally sheet, much as jour-
nal items are posted to a ledger. Each entry is sepa-
rated from adjacent ones by a dash. Each line is then
totaled, and the result entered under the five cent
column of ''Frequency Classes.'' A series of values
thus arranged according to magnitude is known as an
array.

An inspection of the five cent frequencies shows that
we have discovered only a very rough approximation

TABLE II

WAGES IN A CONNECTICUT WOOLEN MILL, JULY, 1891

Frequency Classes—Intervals of: 5¢, 15¢, 25¢, 50¢

| Daily Wage $ | Tally (No. of Workers) | 5¢ | 15¢ | 25¢ | 50¢ | Σ | Daily Wage $ | Tally (No. of Workers) | 5¢ | 15¢ | 25¢ | 50¢ | Σ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| .40 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| .45 |  |  |  |  |  |  |  | Carried over . | 349 | 351 | 349 | 349 |  |
| .50 | 5- | 5 | 5 |  |  | 5 | 2.50 | 1-1- | 2 |  |  |  | 351 |
| .55 |  |  |  |  |  | 5 | 2.55 |  |  |  |  |  |  |
| .60 |  |  |  | 5 |  | 5 | 2.60 |  |  | 0 | 2 |  |  |
| .65 |  |  | 0 |  |  | 5 | 2.65 |  |  |  |  |  |  |
| .70 |  |  |  |  | 42 | 5 | 2.70 |  |  |  |  | 6 |  |
| .75 | 1-9- | 10 |  |  |  | 15 | 2.75 | 1-1-1-1- | 4 | 4 |  |  | 355 |
| .80 | 2-14- | 16 | 29 |  |  | 31 | 2.80 |  |  |  | 4 |  |  |
| .85 | 1-2- | 3 |  | 37 |  | 34 | 2.85 |  |  | 0 |  |  |  |
| .90 | 1-3-4- | 8 |  |  |  | 42 | 2.90 |  |  |  |  |  |  |
| .95 |  |  | 10 |  |  | 42 | 2.95 |  |  |  |  |  |  |
| 1.00 | 1-1- | 2 |  |  |  | 44 | 3.00 | 3- | 3 | 3 |  |  | 358 |
| 1.05 | 2- | 2 |  |  |  | 46 | 3.05 |  |  |  | 3 |  |  |
| 1.10 | 14-1-4- | 19 | 38 | 58 |  | 65 | 3.10 |  |  |  |  |  |  |
| 1.15 | 1-1-3-12- | 17 |  |  | 117 | 82 | 3.15 |  |  | 0 |  | 3 |  |
| 1.20 | 1-17- | 18 |  |  |  | 100 | 3.20 |  |  |  |  |  |  |
| 1.25 | 2-1-13-3-2-4-3- | 28 | 56 |  |  | 128 | 3.25 |  |  |  |  |  | 358 |
| 1.30 | 7-1-2- | 10 |  | 59 |  | 138 | 3.30 |  |  | 0 | 0 |  |  |
| 1.35 | 8-1-8-2- | 19 |  |  |  | 157 | 3.35 |  |  |  |  |  |  |
| 1.40 | 1-1- | 2 | 21 |  |  | 159 | 3.40 |  |  |  |  |  |  |
| 1.45 |  |  |  |  |  | 159 | 3.45 |  |  |  |  |  |  |
| 1.50 | 2-3-1-1-4- | 11 |  |  |  | 170 | 3.50 | 1- | 1 | 1 |  |  | 359 |
| 1.55 |  |  | 14 |  |  | 170 | 3.55 |  |  |  | 1 |  |  |
| 1.60 | 2-1- | 3 |  | 106 |  | 173 | 3.60 |  |  | 0 |  |  |  |
| 1.65 | 1- | 1 |  |  |  | 174 | 3.65 |  |  |  |  | 1 |  |
| 1.70 | 2-89- | 91 | 159 |  | 180 | 265 | 3.70 |  |  |  |  |  | 359 |
| 1.75 | 1-3-2-60-1- | 67 |  |  |  | 332 | 3.75 |  |  | 0 | 0 |  |  |
| 1.80 |  |  |  |  |  | 332 | 3.80 |  |  |  |  |  |  |
| 1.85 | 1-1- | 2 | 7 | 74 |  | 334 | 3.85 |  |  | 0 |  |  |  |
| 1.90 | 1-4- | 5 |  |  |  | 339 | 3.90 |  |  |  |  |  |  |
| 1.95 |  |  |  |  |  | 339 | 3.95 |  |  | 1 |  |  |  |
| 2.00 | 1-1- | 2 | 2 |  |  | 341 | 4.00 | 1- | 1 |  |  |  | 360 |
| 2.05 |  |  |  |  |  | 341 | 4.05 |  |  | 0 | 1 |  |  |
| 2.10 | 2-1- | 3 |  | 9 |  | 344 | 4.10 |  |  |  |  |  |  |
| 2.15 |  |  | 7 |  |  | 344 | 4.15 |  |  |  |  | 2 |  |
| 2.20 | 4- | 4 |  |  | 10 | 348 | 4.20 |  |  |  |  |  |  |
| 2.25 | 1- | 1 |  |  |  | 349 | 4.25 | 1- | 1 | 1 |  |  | 361 |
| 2.30 |  |  | 1 |  |  | 349 |  |  |  |  |  |  |  |
| 2.35 |  |  |  | 1 |  | 349 |  |  |  |  |  |  |  |
| 2.40 |  |  |  |  |  | 349 |  |  |  |  |  |  |  |
| 2.45 |  |  | 2 |  |  | 349 |  |  |  |  |  |  |  |
| Totals ...... |  | 349 | 351 | 349 | 349 |  | Totals ...... |  | 361 | 361 | 361 | 361 |  |

to the theoretical frequency curve. We therefore experiment with larger groupings to see if we can thus obtain more distinctive results. Classes at fifteen, twenty-five, and fifty cent intervals are shown in the designated columns. These classes are found by adding the five cent frequencies falling within the limits indicated by the horizontal bars. Obviously,

several variations of these groupings could be made
by beginning at different points in the scale; but the
arrangement here chosen, which gives regular classes
back to the zero point, is the most natural one to take.
Comparing the different groupings, we see that the
twenty-five and fifty cent intervals give results as
smooth as we are likely to get. Since a classification
with larger intervals promises to be too indefinite, it
is useless to carry our frequency classifications
further.

**The Derived or Special Purpose Table.** In order to
give a summarized presentation of the fifty cent fre-
quencies, Table III has been drawn up. In this table
the classes are designated by stating the upper and
lower limits, as $.50 to $.95, inclusive. If the class
interval is so arranged that the mid-point falls at a
round number, the class may be designated by this
number. Thus the twenty-five cent frequencies could

TABLE III
WAGES IN A CONNECTICUT WOOLEN MILL *

| WAGE PER DAY DOLLARS | NUMBER AND PERCENTAGE OF WORKERS DISTRIBUTED ACCORDING TO DAILY WAGES, JULY— | | | | | |
|---|---|---|---|---|---|---|
| | 1870 | | 1880 | | 1891 | |
| | NO. | % | NO. | % | NO. | % |
| 0 to .45 | 2 | 1.8 | 2 | 0.9 | 0 | 0 |
| .50 to .95 | 18 | 16.7 | 58 | 27.2 | 42 | 11.6 |
| 1.00 to 1.45 | 54 | 50.0 | 126 | 59.2 | 117 | 32.4 |
| 1.50 to 1.95 | 22 | 20.4 | 17 | 8.0 | 180 | 49.8 |
| 2.00 to 2.45 | 3 | 2.8 | 1 | 0.5 | 10 | 2.8 |
| 2.50 to 2.95 | 6 | 5.6 | 3 | 1.4 | 6 | 1.7 |
| 3.00 to 3.45 | 1 | 0.9 | 6 | 2.8 | 3 | 0.8 |
| 3.50 to 3.95 | 2 | 1.8 | 0 | 0 | 1 | 0.3 |
| 4.00 to 4.45 | 0 | 0 | 0 | 0 | 2 | 0.6 |
| Total | 108 | 100 | 213 | 100 | 361 | 100 |

* Aldrich Report, pp. 1463 ff.

be tabulated according to the nearest quarter dollar.
The usual method is, however, to state the limits, as
here shown.

In contrast with Table I, which presented in orderly
form practically all the detailed data of our study, this
table is a derived or special purpose table. It aims to
present only certain features of the wage schedules,
and therefore purposely omits details. It is in the
nature of a generalization, condensing the original
facts into as brief a compass as is practicable.

The preparation of such a table usually calls for
both consideration and skill. The bracketing system
used in the headings, or captions, is obvious—the wider
blocks bracket and designate the smaller ones imme-
diately beneath. Which set of subdivisions are entered
as captions, and which are entered in the stub to the
left, is usually determined by the exigencies of space.
The arrangement of the details will depend upon the
nature of the table. In census tables, for example,
the current date is placed in the first column, to the
left, because of its greater importance, thus reversing
the chronological order. In such tables, also, totals
will be given the place of prominence at the top,
directly beneath the caption. In an elaborate table
percentage columns will be placed together, or in a
separate table, to allow of easy comparison. Correla-
tive items in the caption or stub should be arranged
in some logical order, whether by magnitude as in the
case of the frequency classes, chronologically as the
successive wage distributions, geographically as in the
case of a census list of states, by order of origin, or
merely alphabetically.

In computing the percentage columns, each frequency is divided by the total. The work will be sufficiently accurate if computed on a slide rule or string chart. The percentage totals will not necessarily come to exactly one hundred per cent, because of the inaccuracies involved in cutting off decimals. If it is desired, however, they may be brought to the proper
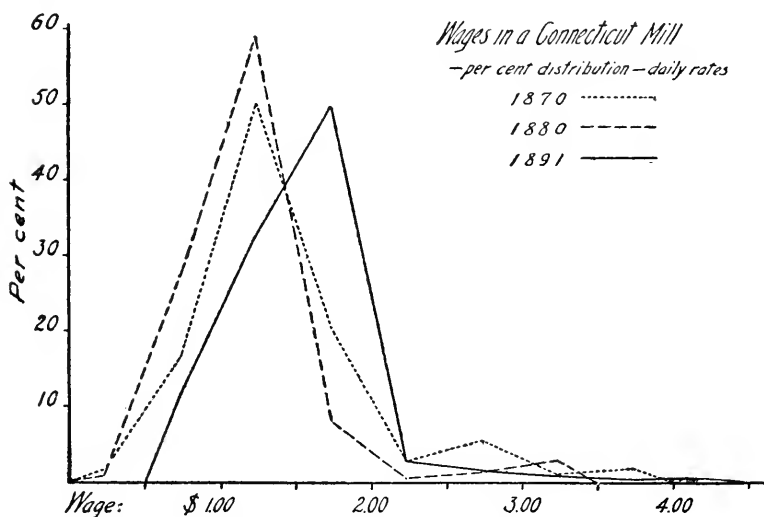


Figure 2.   Frequency polygons

total merely by turning one or two items that stand at or near five in the first decimal dropped. Thus 7.55 may be written 7.6 or 7.5, according to which is needed to make up the total of one hundred; or 7.56 might even be written 7.5 if no better can be done. A sufficient degree of accuracy can always be secured by extending the number of decimals retained. But, in general, the special purpose table should not exhibit meticulous accuracy. Decimals should be shortened

or dropped, fractions avoided, and large numbers rounded.

**The Frequency Polygon.** The percentage columns of Table III may now be graphically presented by "frequency polygons," as shown in Figure 2. The percentage columns are here used in preference to the absolute numbers because they reduce the three polygons to the same scale. It will be seen that in drawing the frequency polygons the points representing the percentage frequencies are plotted directly above the mid-point of each class, respectively. These points are then joined, the individual years being distinguished by different kinds of lines. The graph brings out very well the general advance in minimum, maximum, and mean wages that occurred in 1891 in the mill under consideration. The data for a single year may also be graphed as a "rectangular histogram," as illustrated in the next chapter (Fig. 3, p. 25; solid line). Both the frequency polygon and the rectangular histogram are frequency curves approximately expressed.

**Difficult Features of Tabulation.** Before leaving the subject of tabulation, a few general suggestions may be made. In the tabulation considered in this chapter, the problem of interpreting the term "wages" has been solved for us by the Aldrich Report. But if we had undertaken the task of filling in the original schedules by actual field work, we should have been faced with the difficulty of drawing a somewhat artificial line distinguishing wages from salaries, and perhaps from commissions and other direct or indirect income. From the standpoint of economic theory, of course, salaries are generically wages. But in practice

payments for relatively responsible and skilled work, contracted usually on the basis of a considerable period of time, and carrying some degree of stability of tenure, are classed as salaries. They are excluded from wage schedules as being presumably determined less directly by supply and demand considerations. Likewise most statistical units, however precise and simple they may appear at first glance, usually present many difficulties when they are applied to real conditions. Precisely what, for example, should be included in a tabulation as a book, a farm, an accident, a ton-mile? Almost any unit that may be chosen will be found to call for careful discrimination, and an examination of current usage.

A further difficulty is encountered when data concerning given units are being gathered and compared over a certain period of time, or from different contemporaneous environments. It often happens that the definition of the unit varies at different times or in different places; or perhaps the basis of estimating the frequencies may be altered, or comparisons may be invalidated by changing conditions. In our study of the Connecticut mill we may evidently assume that the basis of the classification has not changed materially through the period studied, since the occupations classed as wage-earning are specified. But we might modify our interpretation of the change in wage levels upon observing that processes of work had altered, that the work-day was shortening, that child labor legislation was affecting the personnel, that the percentage of female workers shifted from 28% to 19%, and then to 32%, or that the cost of living had fallen. Thus it is always necessary in comparative studies to

consider carefully both the environmental conditions and the statistical units employed.

The statistician who is at all ingenious will discover many short cuts to lighten the work of tabulation. A method which is often useful is that of entering the original data on $3 \times 5$ or $4 \times 6$ cards. Suppose, for example, that we wished to classify the students of a given college according to their entrance grades, the class of school from which entering, fraternity membership, and scholastic standing during their college course. A numbered card could be prepared for each student, and the desired data entered. The cards could then be sorted as desired, the sub-totals could be determined, or the data listed. If, however, such work is to be done on an extended scale, a tabulating machine will be required. Such a machine automatically sorts and counts special cards on which the required data have been recorded by a keyed punch. The larger business houses are using machine tabulators increasingly; and of course extensive compilations like a census are prepared principally by machines.

**Library Work.** The subject of tabulation has been extensively treated by writers on statistics. Day's article, cited below, will prove to be very valuable to the student. It may be found reprinted in Secrist's "Readings," together with another excellent article on the same subject. Chapter IV of the same book is especially pertinent to the subject of wage tabulations. Chapter VII of Rugg's text-book presents a concise description of the frequency curve. Machine tabulation is described in the circulars of the Tabulating Machine Company, of New York.

*REFERENCES*

Bailey and Cummings, *Statistics,* Chapters I-V.

Bowley, Arthur L., *Elements of Statistics,* Chapter IV.

Day, E. E., "Standardization of the Construction of Statistical Tables," *Quarterly Publications of the American Statistical Association,* March, 1920, pp. 59-66.

Koren, John, *A History of Statistics.*

Rugg, H. O., *Statistical Methods Applied to Education,* Chapter VII.

Secrist, Horace, *An Introduction to Statistical Methods,* Chapters I-V.

Secrist, Horace, *Readings and Problems in Statistical Methods,* Chapters I-V.

Yule, G. U., *An Introduction to the Theory of Statistics,* Chapter VI.

## EXERCISES [1]

1. Draw up tally sheets and frequency classifications for the years 1870 and 1880.

2. Tabulate the twenty-five cent classes, showing both absolute and percentage figures, for the years 1870, 1880, and 1891. Draw frequency polygons from the percentage data.

3. Similarly tabulate and graph the fifteen cent classes.

4. Graph the five cent classes for the years 1870, 1880, and 1891.

5. Draw rectangular histograms of the fifty cent frequencies for 1870 and 1880. Compare the relative advantages of the frequency polygon and the rectangular histogram.

6. Classify and tabulate separately the female workers for the years 1870, 1880, and 1891. (Starred items—see footnote, Table I.)

7. Obtaining data from the Aldrich Report, study the wages paid in Establishment No. 86 in July of 1875 and 1885. Classify, tabulate, and graph as for the other years studied.

8. Obtain the average scholarship grades for a selected group of students (100 or more), classify these grades, tabulate, and draw a frequency polygon.

9. Toss two coins twenty-five times, keeping a record of the number of heads thrown at each toss. Classify and tabulate the results, and draw a frequency polygon. Toss four coins fifty times, making similar records. What principle is illustrated?

[1] Before beginning a notebook the student should read Appendix I.

## TYPES AND MEASURES OF DISPERSION

After the frequency distribution of a given array
has been presented in suitable form, there remains the
task of finding simple numerical measures by which
it may be summed up for purposes of ready descrip-
tion and comparison. The two features thus to be
expressed are the typical wage and the degree of dis-
persion or ''spread'' about the type. As a wage type,
and a base from which dispersion may be measured,
the common average, or arithmetic mean, will immedi-
ately suggest itself. In the case of the wage rolls given
in the preceding chapter, the average may be most con-
veniently found by multiplying the wages, as tabulated
at five cent intervals, by their respective frequencies.
The sum of these products, divided by the number of
workers, is the average. It is the weighted average of
the class values at five cent intervals, since these values
are emphasized in accordance with their frequencies.
We shall find that weighted averages are sometimes
taken in which the weights are derived estimates of
the importance which should be attached to the values,
respectively; but in this case the weighting amounts
simply to a summing up of the original wages. The
formula for the weighted average is $\dfrac{\Sigma FV}{N}$ (summa-
tion of the frequencies times the class values, divided

by the number of items). The accompanying table
(Table IV), derived from Table II, shows the process
of finding the average wage for the year 1891.

TABLE IV

WAGE ROLL AND AVERAGE WAGE

CONNECTICUT MILL, JULY, 1891

| Single Wage | No. of Workers | Total Wage |
|---|---|---|
| $ .50 | 5 | $2.50 |
| .75 | 10 | 7.50 |
| .80 | 16 | 12.80 |
| .85 | 3 | 2.55 |
| .90 | 8 | 7.20 |
| 1.00 | 2 | 2.00 |
| 1.05 | 2 | 2.10 |
| 1.10 | 19 | 20.90 |
| 1.15 | 17 | 19.55 |
| 1.20 | 18 | 21.60 |
| 1.25 | 28 | 35.00 |
| 1.30 | 10 | 13.00 |
| 1.35 | 19 | 25.65 |
| 1.40 | 2 | 2.80 |
| 1.50 | 11 | 16.50 |
| 1.60 | 3 | 4.80 |
| 1.65 | 1 | 1.65 |
| 1.70 | 91 | 154.70 |
| 1.75 | 67 | 117.25 |
| 1.85 | 2 | 3.70 |
| 1.90 | 5 | 9.50 |
| 2.00 | 2 | 4.00 |
| 2.10 | 3 | 6.30 |
| 2.20 | 4 | 8.80 |
| 2.25 | 1 | 2.25 |
| 2.50 | 2 | 5.00 |
| 2.75 | 4 | 11.00 |
| 3.00 | 3 | 9.00 |
| 3.50 | 1 | 3.50 |
| 4.00 | 1 | 4.00 |
| 4.25 | 1 | 4.25 |
| Total...... | 361 | $541.35 |
| Average | | 1.49958 $1.50 |

**The Mode.** In addition to the arithmetic mean, there
are other types which the statistician uses in summar-

izing an array and measuring dispersion. One of these is the mode. The mode is applicable, however, only to frequency distributions which conform in their general outlines to the theoretical frequency curve. It is the value which lies at the point of greatest frequency, and in the normal curve is therefore identical with the average. In the graph of a frequency distribution it is easily recognizable as the value indicated on the horizontal scale at the point directly under the highest point of the curve. (See Figure 3, page 25.)

The mode is particularly useful in connection with those frequency curves which, though conforming in general outlines to the theoretical, are extended more on the one side than the other. Such curves are said to be skewed. The wage data studied in the preceding chapter gives curves which are somewhat skewed to the right, so that a small secondary mode sometimes appears. But in their original five cent frequencies they do not give a smooth enough curve to allow of a very definite mode. When, however, such curves are strongly skewed, and are yet passably smooth, the mode is preferable to the average as a type of the array.[1] Suppose, for example, that in a wage array a few very large salaries are included. In such a case the average may fall between the wages and the salaries, at a point where the frequencies are small. The mode, on the other hand, states the wage or salary most frequently paid. It is not affected by the skewed extreme of the curve; that is, by the relatively small number of large salaries.

---

[1] Frequency curves that are strongly skewed to the right will sometimes appear normal if transferred to semi-logarithmic paper, the

**Determining the Mode.** In frequency distributions that are irregular the mode may often be approximately determined by a study of the larger frequency groupings. Let us take as an illustration the twenty-five cent frequency classes derived from the 1891 wage data. These are shown in the third column of Table V, the first and second columns being the data from which they are derived, as given in Table II. The latter part of the series is omitted, however, since it cannot affect the position of the mode. In the fourth and succeeding columns, variations of the twenty-five cent frequencies are formed by beginning the classes at different points in the scale. In each case the mode should lie somewhere within the class having the largest frequency; that is, it should lie in the following classes:

$1.50 — $1.70
1.55 — 1.75
1.60 — 1.80
1.65 — 1.85
1.70 — 1.90

Since the only point common to these five classes is $1.70, this sum may be regarded as the mode. It will be seen, however, that this is the same value which would be taken as the mode on the basis of the five cent classes. The same result would in this case also be obtained by using the fifty cent classes. Ordinarily, this method is applied only to the largest classes which it is practicable to use in the frequency classification. It may give a quite different result from that which is obtained from the smallest classes.

logarithmic scale being used as the base line. When this is the case, the distribution is normal on the basis of the geometric rather than the arithmetic mean (see page 94).

TABLE V

THE MODE

WAGE ROLL, CONNECTICUT MILL, JULY, 1891

| V | F | FREQUENCIES IN 25c CLASSES | | | | | SUMMARY |
|---|---|---|---|---|---|---|---|
| | | VARIOUS GROUPINGS | | | | | |
| $ .40 | 0 | | 5 | | | | 5 |
| .45 | 0 | | | 5 | | | 5 |
| .50 | 5 | | | | 5 | | 5 |
| .55 | 0 | | | | | 5 | 5 |
| .60 | 0 | 5 | | | | | 5 |
| .65 | 0 | | 10 | | | | 10 |
| .70 | 0 | | | 26 | | | 26 |
| .75 | 10 | | | | 29 | | 29 |
| .80 | 16 | | | | | 37 | 37 |
| .85 | 3 | 37 | | | | | 37 |
| .90 | 8 | | 29 | | | | 29 |
| .95 | 0 | | | 15 | | | 15 |
| 1.00 | 2 | | | | 31 | | 31 |
| 1.05 | 2 | | | | | 40 | 40 |
| 1.10 | 19 | 58 | | | | | 58 |
| 1.15 | 17 | | 84 | | | | 84 |
| 1.20 | 18 | | | 92 | | | 92 |
| 1.25 | 28 | | | | 92 | | 92 |
| 1.30 | 10 | | | | | 77 | 77 |
| 1.35 | 19 | 59 | | | | | 59 |
| 1.40 | 2 | | 42 | | | | 42 |
| 1.45 | 0 | | | 32 | | | 32 |
| 1.50 | 11 | | | | 16 | | 16 |
| 1.55 | 0 | | | | | 15 | 15 |
| 1.60 | 3 | 106M | | | | | 106 |
| 1.65 | 1 | | 162M | | | | 162 |
| 1.70 | 91 | | | 162M | | | 162M |
| 1.75 | 67 | | | | 161M | | 161 |
| 1.80 | 0 | | | | | 165M | 165 |
| 1.85 | 2 | 74 | | | | | 74 |
| 1.90 | 5 | | 9 | | | | 9 |
| 1.95 | 0 | | | 9 | | | 9 |
| 2.00 | 2 | | | | 10 | | 10 |
| 2.05 | 0 | | | | | 5 | 5 |
| 2.10 | 3 | 9 | | | | | 9 |
| 2.15 | 0 | | 8 | | | | 8 |
| 2.20 | 4 | | | 8 | | | 8 |
| 2.25 | 1 | | | | 5 | | 5 |
| 2.30 | 0 | | | | | 5 | 5 |
| 2.35 | 0 | 1 | | | | | 1 |
| 2.40 | 0 | | | | | | |
| 2.45 | 0 | | | | | | |
| etc. | | | | | | | |

As applied to the given wages, however, the foregoing method of locating the mode is objectionable. The $1.70 frequency is relatively so large that in each classification it determines the mode without allowing due weight to the large frequencies a little further down the scale. The latter might be considered a secondary mode, but we shall here assume that the use of more extensive data would result in a single mode. We may therefore illustrate the use of a method which
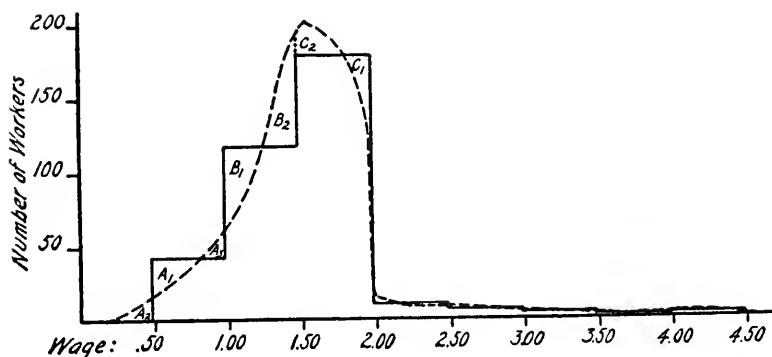


Figure 3. Rectangular histogram of wages in a Connecticut mill, 1891, fifty cent classes (solid line), and smoothed frequencies (broken line).

is applicable to irregular frequencies, or to data presented in only a few large classes. It will, of course, be understood that with such limited data no very dependable result can be obtained.

In using this method, we shall not only approximately determine the mode, but draw the smoothed curve as well. The method is illustrated in Figure 3. The frequencies are first represented by a rectangular histogram. In drawing the histogram the vertical lines should theoretically be drawn at a point

midway between the two adjacent class limits; for example, the line separating the first and second classes falls at .475.[1] It is evident that the mode lies in the class \$1.50 to \$1.95, but it is desirable to locate it somewhat precisely within the class. This may be done by dividing the class interval into two parts proportional inversely to the adjacent frequencies.[2] In so doing the class limits are considered \$1.475 and \$1.975, as drawn. The division may readily be constructed geometrically, or it may be computed as follows:

$$\$1.475 + \frac{10}{117 + 10} \times \$.50 = \$1.51$$

The formula for this operation is,

$$M = L_1 + \frac{F_n}{F_m + F_n} \times C$$

in which

$\quad\quad\quad\quad M =$ the mode.

$\quad\quad\quad\quad L_1 =$ the lower limit of the modal class.

$F_m$ and $F_n =$ frequencies adjacent to the one containing the mode, in the order named.

$\quad\quad\quad\quad C =$ the class interval.

**Smoothing the Frequencies.** After the mode has been determined, the histogram may be smoothed into a frequency curve. This curve is drawn to conform as closely as possible to the theoretical bell-shaped curve; and yet to maintain, frequency by frequency, the same area as the original rectangular figure. Thus in the drawing $A_1 = A_2 + A_3$, $B_1 = B_2$, and $C_1 = C_2$. The curve culminates approximately at the mode as

[1] The class limits should be so placed that the items in the modal class average close to the mid-point of the class.

[2] It is sometimes preferable to take the average of the items in the modal and adjacent classes, or to include more classes.

previously determined, but the height is merely esti-
mated with reference to the required area.    As thus
drawn, the curve presents an estimate of the probable
distribution of the economic values expressed by wage
rolls of the type studied.    The irregularity of the data,
however, makes it far from typical.

**The Median.**    Another type often used to represent
a given array is the median. The median is the value of
the middle item in an array.    The number of this item
is found by the formula, $\dfrac{N + 1\,^{1}}{2}$; and its value may be
determined by reference to the summation column of
the frequency table.    For example, in the 1891 wage
data the median item is number 181, and its value as
determined by means of the summation column is
$1.70.    That is, the 181st item falls within the $1.70
class.    In case the median item should prove to be
fractional, and should fall between two frequencies, the
median would not be precisely determined.    Suppose,
for example, that the median item had been number
174½.    In such a case the median would lie between
the limits $1.65 and $1.70.

**Comparison of the Three Types.**    Theoretically, in a
frequency curve having very small class intervals the
median value is indicated at the foot of a perpendic-
ular line which bisects the area of the curve.    The aver-
age, on the other hand, would lie at the foot of a similar
perpendicular which would balance as an axis the
weight of the two sides, supposing the area of the curve

---

[1] The unit is added to counterbalance the space from 0 to 1.    Or, the
formula may be considered as an expression of the average of the
extreme ordinals of the array.

to have been cut out of a material of uniform weight. When skewness is regular, the mode, median, and average are located on the value scale in the order named, the intervals separating them being in the ratio of about two to one. In the normal curve the three types are identical.[1]

The use of the average, median, and mode as types

SALARIES PAID IN REPRESENTATIVE UNIVERSITIES AND COLLEGES IN THE UNITED STATES IN 1919-20.

*Public institutions*

| TITLE OF POSITION | NUMBER OF PERSONS | MINIMUM SALARY | MAXIMUM SALARY |
|---|---|---|---|
| President or chancellor......... | 77 | $2,500 | $12,500 |
| Dean or director............... | 367 | 1,200 | 10,000 |
| Professor...................... | 2,460 | 300 | 10,000 |
| Associate professor............. | 822 | 300 | 4,000 |
| Assistant professor............. | 1,705 | 500 | 4,000 |
| Instructor..................... | 2,138 | 300 | 3,100 |
| Assistant...................... | 855 | 75 | 2,500 |

| | AVERAGE SALARY | MEDIAN SALARY | MOST FREQUENT SALARY |
|---|---|---|---|
| President or chancellor......... | $6,647 | $6,000 | $6,000 |
| Dean or director............... | 3,819 | 3,500 | 3,000 |
| Professor...................... | 3,126 | 3,000 | 3,000 |
| Associate professor............. | 2,514 | 2,500 | 3,000 |
| Assistant professor............. | 2,053 | 2,000 | 1,800 |
| Instructor..................... | 1,552 | 1,500 | 1,500 |
| Assistant...................... | 801 | 750 | 1,200 |

[1] Two other forms of the average are sometimes used in statistical work. One is the geometric mean. This may be found by averaging the logarithms of the numbers instead of the numbers themselves. It is the $n$th root of the product of the numbers. Some statisticians advocate the use of the geometric mean in finding the average periodic change in prices. Considered merely as an average of prices apart from the use of weights, the geometric mean is logically correct because it measures ratios of divergence rather than absolute amounts. The other type of average is the harmonic mean, which has occasionally been applied to the same purpose. For two numbers, $a$ and $b$, it is computed by the formula $\frac{2\,a\,b}{a+b}$. In arithmetic this is the formula which is used to find an average rate of travel when two rates for two equal distances are given. In general, it may be described as the reciprocal of the average of the reciprocals of the given numbers.

may be illustrated by the foregoing table which was issued by the United States Bureau of Education and reprinted in the *Monthly Labor Review* of January, 1921. In this table the term "most frequent salary" signifies the mode.

**Quartile Deviation.** The types we have now considered are used as the basis for measuring dispersion, though the average is more commonly employed than the other two. The simplest measure of dispersion is related to the median, and is called the quartile deviation. This is found by computing the value of the first and third quartiles of an array. The quartiles are analogous to and include the median, being located at the quarter divisions of the array. The location of the first quartile is found by the formula $\frac{n+1}{4}$, and of the third quartile by the formula $\frac{3(n+1)}{4}$. The values of these items are determined by reference to the frequency table in the same manner as the median value was determined. The second quartile is, of course, identical with the median. The quartile range is found by subtracting the value of the first quartile from the value of the third, and the quartile deviation is half of this difference. It may be seen by reference to Figure 1 that the quartile deviation as thus found is simply the average distance between the median and the adjacent quartiles, as measured on the base line. In the 1891 wage roll the first quartile is item No. 90½, and its value is $1.20. The third quartile is No. 271½, and its value is $1.75. The quartile range is therefore $.55, and the quartile deviation is $.28. This means

that half the workers receive wages that fall within a range averaging twenty-eight cents above and below the median; that is, between \$1.20 and \$1.75.[1]

For purposes of comparison the quartile deviation should usually be reduced to a percentage basis. This is done by dividing it by a value regarded as typical of the array. Since the quartile deviation is related to the median, it would appear logical to take this type as a base. But it is customary to take instead a point lying midway between the first and third quartile values; that is, the average of the two. In a perfectly regular curve this value would naturally be identical with the median. The reason for taking this base is obviously that it is the point from which the quartile deviation is assumed to be directly measured. The formula for it is $\dfrac{Q_3 + Q_1}{2}$ (third quartile plus first quartile, divided by two). The formula for the quartile deviation is $\dfrac{Q_3 - Q_1}{2}$. The latter divided by the former is $\dfrac{Q_3 - Q_1}{Q_3 + Q_1}$, which is therefore the formula for the coefficient of quartile deviation.

**Interpolation.** Before leaving the quartiles, a method of locating them by interpolation between the class intervals should be described. We will illustrate the method by applying it to the 1891 wage data, though

[1] The quartile deviation is also called the ''probable error'' of a frequency distribution. The term is derived from the ''Theory of Errors,'' and connotes the central range of the distribution within which an item added to the series will have an even chance of falling.

in fact the five cent classes give quartile values precise enough for most purposes. The type of problem to which the method is best adapted is one in which an array as given is classified only in a few large groups. But supposing that it is desirable to know the quartile values very precisely in the 1891 wage data, we may find them as described below.

In assuming that we may interpolate at any point between the items of the original frequency classification, we are evidently regarding the series as continuous rather than discrete. In the case of the wage roll we shall be dealing, then, with the theoretical economic values underlying the wages as paid. The actual frequencies are therefore to be considered as indicating proportionate numbers, which may be increased indefinitely as in the case of any multiple ratio. The original discrete five cent classes are now to be considered as having continuous class intervals; for example, a fifty cent wage is taken to indicate an economic value within the limits $.475 and $.525. ' The frequencies are assumed to be equally spaced between these limits.

When about to interpolate, we locate the first quartile by the formula $\dfrac{N}{4}$, instead of $\dfrac{N+1}{4}$ as in the previous case. The second and third quartiles are two and three times this number, respectively. The reason for omitting the unit in the formula, and for ignoring it also in an analogous formula for sub-dividing the class, is that our hypothesis of a continuous scale renders the unit of negligible value. It is as if

the items were regarded as being groups of thousands or millions; that is, as if they were indefinitely subdivisible. Having located the quartile items, we find the class in which they fall by inspection of the frequency table, as before. We next find the position of the quartile within the class; that is, the fraction of the interval that it is advanced beyond the lower limit. The corresponding value is then determined. The process is the same as that used in interpolating in logarithmic or other tables.

In the 1891 wage data, the first quartile is located at item $90\frac{1}{4}$. This item falls in the class having theoretically a lower limit of \$1.175 and an upper limit of \$1.225. The preceding class ends with the 82nd item, and the quartile is therefore advanced $8\frac{1}{4}$ items in its own class of 18 items. This advance is $8\frac{1}{4} \div 18$, or .46 of the class interval. This fraction of the class interval of \$.05, is \$.023, which, added to the lower limit of the class, gives the quartile value of \$1.198. Read to the nearest cent, this value happens to be the same as that obtained without interpolation.

The process may be summed up as follows:

$$Q = L_1 + \frac{I}{F}. C$$

in which,

 $Q$ = Quartile value
 $L_1$ = Lower limit of class containing quartile
 $I$ = Quartile item minus last item of preceding class [1]
 $F$ = Frequency of class containing quartile
 $C$ = Interval of same class

---

[1] "Item" here refers to the number, not the value.

**Quartile Dispersion.**  A statement of the quartiles and the highest and lowest wage paid serves to give a fairly good idea of a frequency distribution even without any further computation of precise measures of dispersion.  In Table VI such a statement is presented

RANGE AND TREND OF WAGES, 1870-1891, IN A CONNECTICUT WOOLEN MILL
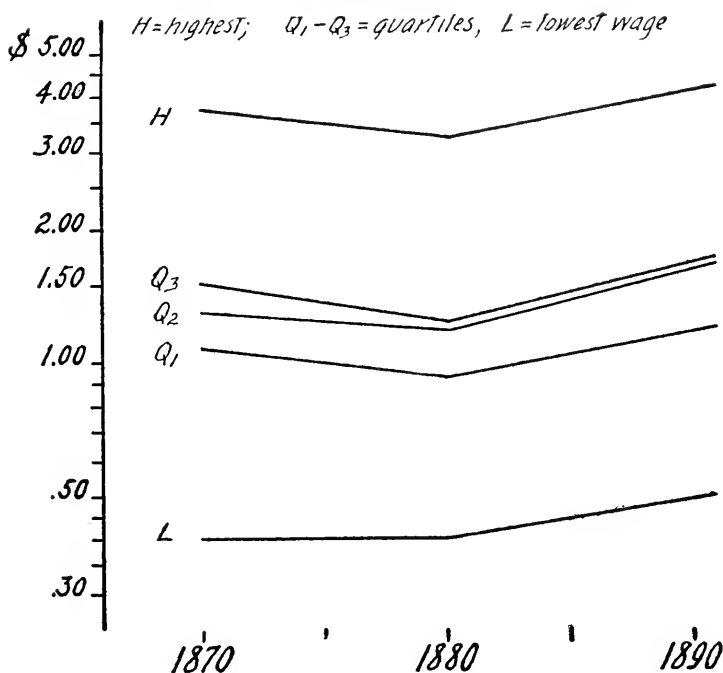


Figure 4.  Semi-logarithmic, or ratio paper

for the wage data of 1870, 1880, and 1891, together with the quartile deviations and their coefficients.  The table includes the interpolated values, although, as has been intimated, their computation is hardly worth while here except as an illustration of the method.  In Figure 4 the discrete quartiles and limits are shown

graphed upon semi-logarithmic paper. The vertical
scale of this paper is similar to the scale of a slide-
rule, hence the ratio of periodic change may be com-
pared by means of the slant of the lines connecting
the values. To be complete, however, such graphic

TABLE VI

RANGE OF DAILY WAGES IN A CONNECTICUT MILL—SCALE
TAKEN BOTH AS DISCRETE AND CONTINUOUS

| WAGE ARRAY | 1870 | | 1880 | | 1891 | |
|---|---|---|---|---|---|---|
| | DIS-CRETE | CONTINU-OUS | DIS-CRETE | CONTINU-OUS | DIS-CRETE | CONTINU-OUS |
| Lower Limit......... | $ .40 | $ .38 | $ .40 | $ .38 | $ .50 | $ .48 |
| 1st Quartile......... | 1.10 | 1.09 | .95 | .93 | 1.20 | 1.20 |
| 2nd Quartile......... | 1.30 | 1.30 | 1.20 | 1.19 | 1.70 | 1.68 |
| 3rd Quartile......... | 1.50 | 1.51 | 1.25 | 1.24 | 1.75 | 1.73 |
| Higher Limit........ | 3.75 | 3.78 | 3.25 | 3.28 | 4.25 | 4.28 |
| Q. Deviation........ | .20 | .21 | .15 | .16 | .28 | .27 |
| —coefficient...... | 15% | 16% | 14% | 14% | 19% | 18% |

representation should show at least annual data. It
might also very well show decile points; that is, the
wages occurring at the tenths instead of the quarters
of each array.

**The Ogive.** A convenient method of presenting an
array and at the same time of graphically determining
the quartiles, is shown in Figure 5. The construction
is based upon the assumption of a continuous series of
values, and parallels the procedure of finding the quar-
tile values by interpolation. The frequencies are
plotted from the summation column of the original
five cent classes. For convenience of interpretation, a
dot marks the entry as it would be made on the graph
if the series were taken as discrete. A slanting line is
drawn across each class interval, beginning with the
summation total of the preceding class, and ending

with the summation total of the given class.  The given
frequency is thus represented as distributed evenly
through the class.  The resulting figure is known as
an ogive.  To find the quartiles, the vertical scale rep-
resenting the whole array is divided into four equal
parts, and horizontal lines are drawn from the quartile
division points until the ogive is intersected.  From
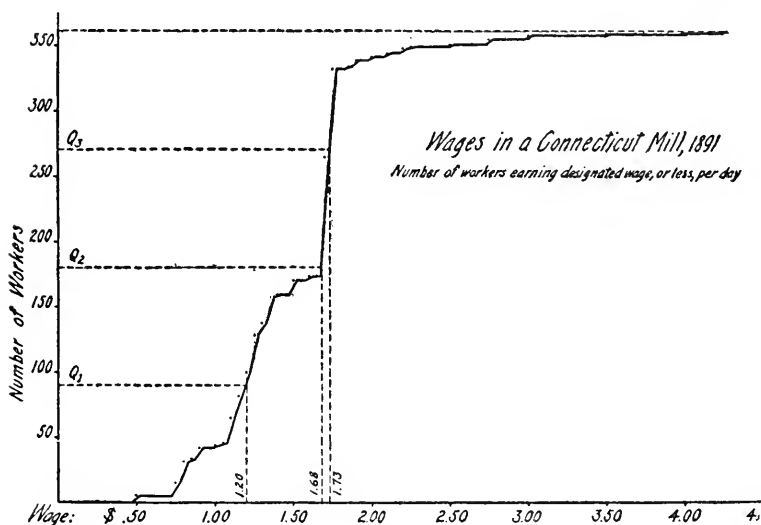


Figure 5.  Cumulative curve, or ogive

the points of intersection perpendiculars are drawn
to the base line.  The foot of each perpendicular marks
upon the horizontal scale one of the quartile values.
The deciles may be found by dividing the horizontal
scale into tenths, and proceeding as before.  This
graphic process, worked out on large sheets of cross-
section paper, is usually the most convenient method
of finding the quartiles or deciles.[1]

[1] A more complex form of the ogive has recently been introduced
for testing the regularity of a frequency distribution.  This ogive is

**Average and Standard Deviation.** We shall now consider the more commonly used mathematical measures of dispersion,—the average deviation and the standard deviation. The former is coming to be fairly well known as applied to economic data. The latter is, however, generally favored by the mathematician, but its chief statistical use at present lies in connection with the measurement of correlation, a subject which will be taken up later.

The principles involved in average and standard deviation may best be illustrated by taking a very simple example. Suppose that four workers are employed at daily wages of \$2.00, \$6.00, \$7.00 and \$9.00, respectively. The average wage is \$6.00. The first wage differs from the average by \$4.00, the second is at the average, the third differs by \$1.00, and the fourth differs by \$3.00. The sum of these differences is \$8.00, or an average of \$2.00 for each wage. The average deviation (A. D.) is therefore \$2.00, which may be taken as a measure of the "spread" of the wages.[1] The standard deviation ($\sigma$) is computed by squaring the deviations, averaging the squares, and finding the square root of this result. In each case a coefficient may be found by dividing by the average wage. The computations are written out in the following form:

drawn upon so-called probability paper, and is constructed from the summated percentage frequencies. The vertical scale of the probability paper is so graduated that a normal curve will form a straight line diagonally across the paper. The divergence of a given distribution from normal may be estimated by its departure from a straight line. The paper for this graph, as well as for other statistical work, may be obtained from the Codex Book Company of New York, or from other publishers of statistical material.

[1] The total spread, or range, from the highest to the lowest wage is sometimes given as an inexpert measure of dispersion. But it is of little value because the wage limits are set by single items which have only a haphazard relation to the rest of the array. The average and standard deviations, however, take account of all the items.

| AVERAGE DEVIATION | | STANDARD DEVIATION | | |
|---|---|---|---|---|
| V | D | V | D | D$^2$ |
| $2 | $4 | $2 | $-4 | 16 |
| 6 | 0 | 6 | 0 | 0 |
| 7 | 1 | 7 | 1 | 1 |
| 9 | 3 | 9 | 3 | 9 |
| 4 ) 24 | 4 ) 8 | 4 ) 24 | 0   4 ) 26 | |
| A = 6 | A. D. = 2 | A = 6 | $\sigma^2 = 6.5$ | |
| | Coef. $=\dfrac{2}{6}=33\%$ | | $\sigma = 2.55$   Coef. $=\dfrac{2.55}{6} = 43\%$ | |

A practical application of average deviation may be cited from Dewing, "Corporation Finance," Vol. III. The writer states that the earnings of corporations producing inexpensive necessities, directly consumed, are most regular; while the earnings of corporations producing expensive indirect goods are least regular. He illustrates the two types of corporations by the Diamond Match Company and the American Locomotive Company, computing the average deviations of the net earnings of the two companies. The coefficient in the first case is 7.1%, and in the second case 50%. The dispersion might have been measured in other ways, as by the standard deviation, but the quartile measure would not be applicable to such a small number of unclassified deviations.

The quartile, average, and standard deviations do not give the same results, as they measure progressively larger portions of the frequency curve (see Fig. 1, p. 10). But in regular distributions, a comparison will be the same whichever measure is used to make the comparison. In an irregular distribution which has a few extreme items, the standard deviation will give an exceptionally large result, since the process of squaring the deviations emphasizes these extremes.

**A Short-cut Method.** The work of finding the average or standard deviation is often rather tedious, particularly when the average is expressed as a decimal. In finding the former, however, it is not important that the average should be expressed very precisely, since the slight error involved in cutting short a decimal is minimized in the process of the work. And in finding the standard deviation, a short-cut process may be used. In this process a convenient average is assumed, and a correction is made later. The method of making the correction may be illustrated by the simple wage scale of four items previously used. The average wage, from which the deviations are to be measured, will be assumed to be $7. Needless to say, this assumption is not here advantageous, though it would have been if the average had been, let us say, $6.75. When the deviations are measured from the assumed average of seven, they give an algebraic sum of —4, which results from the fact that the "correction" appears once in each deviation. Hence the algebraic sum of the deviations, divided by the number of items, will give the "correction,"—the term being taken to mean the sum which must be added to the assumed average to make it the exact average. The correction (K) when found is added to the assumed average, and its square is subtracted from the average squared deviation. By so doing, whatever error may have been involved in assuming an average is eliminated. Otherwise, the work is as before. The computation is set down as indicated at the top of the next page.

**Deviation Computed from Frequency Tables.** The illustration that has been considered thus far in the

$$
\begin{array}{ccc}
\text{V} & \text{D} & \text{D}^2 \\
2 & -5 & 25 \\
6 & -1 & 1 \\
\text{A}_x = 7 & 0 & 0 \\
9 & 2 & 4 \\
\underline{\phantom{--}} & \underline{\phantom{--}} & \underline{\phantom{--}}
\end{array}
$$

$$
\begin{array}{cc}
4)\ \overline{-4} & 4)\ \overline{30} \\
\text{K} = -1 & 7.5 \\
\text{A}_x = \phantom{-}7 & \text{K}^2 = 1 \\
& \overline{\phantom{--}} \\
\text{A} = \phantom{-}\overline{6} & \sigma^2 = 6.5 \\
& \sigma\ = 2.55 \\
& \text{Coef.} = \dfrac{2.55}{6} = 43\%
\end{array}
$$

discussion of average and standard deviation, has been simplified by the fact that the array is not classified into frequency groups. When computed from a frequency table, the average and standard deviations require a somewhat more complex process, though in fact the principles involved are precisely those already explained. The one point to be observed is that the class values must be multiplied by their respective frequencies in order that all items may be taken into account. The method is shown in Table VII, where the wage data of 1891 are again taken up.[1] In these computations, the class values are taken at approximately the mid-points of the class intervals. This is the usual procedure, though a small error is introduced by so doing. An exact computation would require that the actual average value of each class, as determined by dividing the total wages of the class by its frequencies, should be substituted.

[1] A complex graph of a frequency distribution, known as the Lorenz curve, may be described in connection with the data of Table VII. This graph is based upon the F and FV columns, as shown under

TABLE VII

AVERAGE DEVIATION AND STANDARD DEVIATION
WAGE ROLL IN A CONNECTICUT MILL, JULY, 1891

AVERAGE DEVIATION

| V | F | FV | D | FD |
|---|---|---|---|---|
| $ .75 | 42 | $31.50 | $ .78 | $32.76 |
| 1.25 | 117 | 146.25 | .28 | 32.76 |
| 1.75 | 180 | 315.00 | .22 | 39.60 |
| 2.25 | 10 | 22.50 | .72 | 7.20 |
| 2.75 | 6 | 16.50 | 1.22 | 7.32 |
| 3.25 | 3 | 9.75 | 1.72 | 5.16 |
| 3.75 | 1 | 3.75 | 2.22 | 2.22 |
| 4.25 | 2 | 8.50 | 2.72 | 5.44 |
| | 361 | ) 553.75 | | 361 ) 132.46 |
| | | A = 1.53 | | A.D. = .367 |

$$\text{Coef.} = \frac{.367}{1.53} = 24\%$$

Average Deviation. The two columns are reduced to percentages and summated, giving the following results:

| Upper limit of class | F ($\Sigma$) | FV ($\Sigma$) |
|---|---|---|
| $1.00 | 11.6% | 5.7% |
| 1.50 | 44.0 | 32.1 |
| 2.00 | 93.9 | 89.0 |
| 2.50 | 96.7 | 93.1 |
| 3.00 | 98.3 | 96.1 |
| 3.50 | 99.2 | 97.8 |
| 4.00 | 99.4 | 98.5 |
| 4.50 | 100.0 | 100.0 |

The two summated columns are then plotted as coördinates, the first on the horizontal scale and the second on the vertical scale. If the wages were all alike, a direct diagonal would result, while disparity of wages registers in the concavity of the line. The use of the five cent classes would give a more accurate representation. The curve has been often used for presenting a comparison of the distribution of wealth or income in different countries.
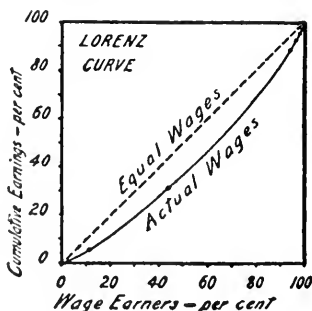
STANDARD DEVIATION (Assumed Average = $1.75) [1]

| V | F | D | FD | FD² |
|---|---|---|---|---|
| $ .75 | 42 | $−1.00 | $−42.00 | $42.00 |
| 1.25 | 117 | −.50 | −58.50 | 29.25 |
| 1.75 | 180 | 0 | 0 | 0 |
| 2.25 | 10 | .50 | 5.00 | 2.50 |
| 2.75 | 6 | 1.00 | 6.00 | 6.00 |
| 3.25 | 3 | 1.50 | 4.50 | 6.75 |
| 3.75 | 1 | 2.00 | 2.00 | 4.00 |
| 4.25 | 2 | 2.50 | 5.00 | 12.50 |
| | 361 | | 361)−78.00 | 361)103.00 |

$$K = -.216 \qquad .2853$$
$$A_x = 1.75 \qquad K^2 = .0467$$

$$A = 1.534 \qquad \sigma^2 = .2386$$
$$\sigma = .49$$
$$\text{Coef.} = \frac{.49}{1.53} = 32\%$$

**Formulas.** The formulas for average and standard deviation are as follows:

A. D. $= \dfrac{\Sigma FD}{N}$ (the deviations here considered positive)

S. D. $(\sigma) = \sqrt{\dfrac{\Sigma FD^2}{N} - K^2}$

in which

F = Frequencies

D = Deviations (from assumed average if followed by a correction)

N = Total number of items in array.

K = Correction for error in assumed average = $\dfrac{\Sigma FD}{N}$

If an average of zero is assumed, the second formula becomes:

S. D. $= \sqrt{\dfrac{\Sigma FV^2}{N} - \left(\dfrac{\Sigma FV}{N}\right)^2}$

[1] A column showing D² is often included, but in most cases it may be omitted and FD² obtained by multiplying D x FD.

In some cases, particularly where a calculating machine is used, this modification of the formula will be found desirable. It calls for the computation of only the columns FV and FV². Its use may be illustrated by reference to Table IV, p. 21. If the first and third columns are multiplied across and totaled, a result of $890.50 will be obtained. This is $\Sigma FV^2$. Dividing by N and subtracting the square of the average, $\left(\dfrac{\Sigma FV}{N}\right)^2$, gives $.218, the square root of which is $.47. This is the standard deviation, obtained somewhat more accurately than before, since smaller class intervals are taken. The modified formula will often be found useful in connection with time series, where no frequencies are involved.

**Summary.** In Table VIII a final summary is made of the dispersion of wages in the Connecticut mill here studied. It is, of course, obvious that the use of a variety of measures is for purposes of illustration only. In practical work of this sort only one measure would be used, probably either the quartile or the average deviation. Skewness would doubtless be compared merely by an inspection of the frequency polygons. Our more exhaustive study, however, gives a very precise picture of the dispersion of wages. On the whole, the "spread" lessens somewhat after 1870, though the change is not great enough to render the data incomparable. It will be noted that the relative skewness of the curves changes but little. Since, then, the dispersion of wages does not materially change, the average wage may be safely taken as an index of the periodic wage level in the given mill.

TABLE VIII

DISPERSION OF WAGES, CONNECTICUT MILL, JULY, 1870, 1880,
AND 1891

| MEASURE | ABSOLUTE | | | RELATIVE | | |
|---|---|---|---|---|---|---|
| | 1870 | 1880 | 1891 | 1870 | 1880 | 1891 |
| Quartile Deviation............ | $ .21 | $ .16 | $ .27 | 16% | 14% | 18% |
| Average Deviation*.......... | .45 | .28 | .37 | 31 | 22 | 24 |
| "        "      (exact).... | .41 | .24 | .35 | 30 | 20 | 23 |
| Standard Deviation*.......... | .62 | .50 | .49 | 44 | 40 | 32 |
| "        "      (exact).... | .61 | .46 | .47 | 44 | 38 | 31 |
| Skewness*................... | .71 | .64 | .55 | 114 | 128 | 113 |

* Computed from fifty cent classes; V = mid-point of class.

**Measurement of Skewness.**  The average deviation,
as we have seen, uses the first power of the deviations,
while the standard deviation uses the second power.
If in an analogous way the third power is used, a
measure of skewness is obtained.  A mathematical
measure of this sort is, however, seldom required in
economic statistics.  The relative degree of skewness
may be roughly determined by comparing the outlines
of frequency curves, or by noting the position of the
modes relative to the averages or medians.  But if an
accurate measure is desired, the following formula
should be employed:

$$\text{Skewness} = \sqrt[3]{\frac{\Sigma FD^3}{N}}$$

The measure may be reduced to a coefficient by divid-
ing by the standard deviation.

**Library Work.**  The subjects of types and disper-
sion are treated in great detail in several of the
standard text-books, such as Bowley's and Yule's.  The

most complete work on the former subject is that of
Zizek, cited below. For an exposition of graphic repre-
sentation, the student should not fail to consult Brin-
ton's work, and the Statistical Atlas of the United
States published in connection with the census. The
ratio chart (semi-logarithmic, or ''arith-log,'' paper)
is well treated in an article by Irving Fisher, as well
as in an article by J. A. Field reprinted in Secrist's
''Readings.'' Whipple's text-book gives an explana-
tion of the use of probability paper as a means of
presenting and testing frequency curves. King's text-
book, page 156, explains the Lorenz curve.

## *REFERENCES*

Bowley, Arthur L., *Elements of Statistics*, Chapters V-VII.
Brinton, W. C., *Graphic Methods for Presenting Facts.*
Fisher, Irving, ''The Ratio Chart,'' *Quarterly Publications
of the American Statistical Association*, June, 1917, pp.
577-601.
King, W. I., *Elements of Statistical Method*, Chapters XII-
XIV.
Marshall, Wm. C., *Graphical Methods*, Chapters I-III.
Secrist, Horace, *Readings and Problems in Statistical Meth-
ods*, pp. 282-305.
Whipple, G. C., *Vital Statistics*, Chapter XII.
Yule, G. U., *An Introduction to the Theory of Statistics*,
Chapters VII and VIII.
Zizek, Franz, *Statistical Averages.*

## EXERCISES

1. Using the five cent frequencies and classes, find the aver-
   age wage for 1870 and 1880.
2. Using 25c class intervals, determine the mode for 1870
   and 1880, following the process illustrated on page 24.
3. Using 50c class intervals, similarly determine the mode
   for 1870 and 1880.
4. Using the 50c frequencies, determine by a mathematical

formula the position of the mode in the 1870 and 1880 data. Draw rectangular histograms and smooth them.

5. Explain why different values for the mode are obtained in the two preceding exercises. Which results are the more valid? Why?

6. Find the quartile items and their values in the 1870 and 1880 wage data by interpolating in the 5c classes. Compute the quartile deviations and coefficients.

7. Draw ogives of the wage data for 1870 and 1880—5c frequencies—showing the quartile values.

8. Summate the percentage frequencies from Table III, page 13, and plot on probability paper.

9. Find the average deviation and coefficient for the 1870 and 1880 wage data, 50c classes.

10. Find the standard deviation and coefficient for the 1870 and 1880 wage data, 50c classes, using the method involving an assumed average and correction.

11. Using data prepared in Exercise 9, draw Lorenz curves of wage distributions in 1870 and 1880. Draw a similar curve from the data of Table IV, page 21.

12. Apply the modified formula for standard deviation (page 41) to the five cent wage data for 1870 and 1880.

13. During a certain period the rate of bank discount was as follows:

| Rate per cent | No. of days |
|---|---|
| $2\frac{1}{2}$ | 174 |
| 3 | 408 |
| $3\frac{1}{2}$ | 132 |
| 4 | 165 |
| $4\frac{1}{2}$ | 36 |
| 5 | 37 |
| $5\frac{1}{2}$ | 20 |
| 6 | 26 |
| 7 | 2 |

(a) Compute the average rate of discount, taking the number of days as the frequencies.

(b) In what classes (rate per cent) do the quartiles fall?

(c) What rate per cent may be taken as the mode? Why?

(No interpolation is required in the above problem.)

14. Find the coefficients of average deviation, standard deviation, and skewness for the following frequency distribu-

tion.  Locate the quartiles by interpolation and check the results by means of an ogive.

| V | F |
|---|---|
| $1 | 1 |
| 2 | 3 |
| 3 | 2 |
| 4 | 2 |
| 5 | 1 |
| 6 | 1 |

15. The following table shows the increase in the cost of living for ten cities from December, 1914, to December, 1920.  Classify these percentages to the nearest multiple of five, and find the average deviation.  (Bureau of Labor Statistics' data.)

| | |
|---|---|
| Boston | 97.4 |
| Buffalo | 101.7 |
| Chicago | 93.3 |
| Cleveland | 104. |
| Detroit | 118.6 |
| Los Angeles | 96.7 |
| New York | 101.4 |
| Philadelphia | 100.7 |
| San Francisco | 85.1 |
| Seattle | 94.1 |

16. Apply the arithmetic formula for determining the mode to the following three time series.  The months may be treated as if they were class intervals, and the percentages as if they were frequencies.  Graph each series and construct a smoothed curve:

Percentage of crops harvested monthly in the United States, as reported by Department of Agriculture.

| Month | Wheat | Corn | Cotton |
|---|---|---|---|
| May | 0.5 | — | — |
| June | 22.0 | 0.1 | — |
| July | 42.3 | 0.1 | 1.4 |
| Aug. | 28.4 | 1.5 | 11.5 |
| Sept. | 6.5 | 15.8 | 31.6 |
| Oct. | 0.3 | 28.3 | 34.4 |
| Nov. | — | 43.3 | 16.0 |
| Dec. | — | 10.9 | 4.7 |
| Jan.-April | — | — | 0.4 |

# CHAPTER III

## INDEXES OF WAGES AND PRICES

**The Nature of Indexes.** A large part of statistical work concerns itself with the making and interpreting of indexes. By an index is meant a number, whether absolute or relative, which is used in comparisons to measure a given condition. Used collectively, the term implies a series of such indexes, forming a multiple ratio. Practically all indexes are compiled by a process of sampling. Thus, though it is impossible to record any large proportion of actual wages and prices, yet it is possible to estimate changes in the wage or price level by the use of well-selected samples. Just what may be regarded as sufficiently complete data in sampling cannot be determined precisely, but must be judged largely on the basis of experience. Straws show which way the wind blows, and likewise the price of a product in a single locality will often accurately reflect the trend of a world market. There is no dependable uniformity, however. Some prices respond quickly and universally to changes in supply or demand, while others move slowly and irregularly. With respect to wages, it is commonly observed that the market is somewhat slow in its movements. In an industrial center like New England, however, the market should be fairly responsive. One might nevertheless hesitate to take the wages at a single mill as an index of wages for the whole country; but com-

parisons will show that such an index would, in fact, have some degree of reliability.

**Indexes of Wages.** The wage averages considered in the preceding chapter will be taken provisionally as indexes of the wage level in the United States. According to these indexes, daily wages in 1870 stood at $1.38, they fell by 1880 to $1.21, but climbed by 1891 to $1.50. The changes may be presented more clearly, however, if the figures are reduced to another form. Since indexes are used as ratios, they may be multiplied or divided through by any factor to suit given requirements. If in this case they are divided by $1.38, the wage in 1870, they are said to be reduced to a base of 1870, since the index for that date will become 100.[1] Expressed literally, the result is 100%, but the per cent sign is usually dropped as being unnecessary in a ratio. The index numbers now read:

| Year. | Wage. |
|-------|-------|
| 1870  | 100   |
| 1880  | 88    |
| 1891  | 109   |

**Indexes of Real Wages.** In a study of changes in the wage level, a further factor of great importance must be taken into consideration. This factor is the cost of living. Changes in the cost of living affect the prosperity of wage earners inversely. Hence "real wages"—a term denoting the purchasing power of wages—will be measured by money wages divided by

---

[1] The base which is theoretically best to use in deriving an index from absolute numbers, is an average of the numbers. Its advantages are, first, that it is a stable value from which to measure the items, and second, that each index is made to suggest its relative position in the series. Applied to the given wage data, the base becomes the average wage of 99c, and the indexes become 101, 89, and 110—each expressing a percentage of the average.

prices.   Whether absolute or relative numbers are taken to measure wages and prices, the quotients may be regarded as comprising an index of real wages, and may be reduced to any desired base.   If wages and prices are expressed as indexes having the same base, then the resulting index of real wages will also have this base.

**Various Wage Indexes.**   The accuracy of our provisional wage index may be tested and the study extended, by the introduction of wage data of a more general character.   These data will be taken from three sources:  (1) ''The Movement of Wages in the Cotton Manufacturing Industry of New England,'' by Professor Stanley E. Howard;  (2) The Aldrich Report, and (3) the publications of the Bureau of Labor Statistics, of the Department of Labor at Washington. The first of these sources gives a carefully prepared index of weekly wages in the Massachusetts cotton manufacturing industry from 1860 to 1914.   It is derived in part from the Aldrich Report, and uses the principles of tabulation and measurement already explained.   The second gives a general index of wages down to 1891, based on wages from many industries, and covering various sections of the country.   The third source furnishes an index of hour rates in the United States from 1840 to 1920.   Hour rates, of course, are not entirely satisfactory as a basis for an index of actual earnings because of the gradual reduction in the length of the working day.   This reduction has, however, been offset by increased over-time pay at higher rates, by a gain in leisure hours, and probably by more regular employment.   And as a matter of fact, the index of hour rates will be found to con-

form somewhat closely to the index of weekly wages. As measuring the cyclical changes in wages, both indexes give practically the same results.

In Table IX the wage indexes from the sources just mentioned are shown for the years 1870, 1880, and 1891, together with the indexes already derived. By means of price indexes—the nature of which will be considered later—nominal wages are reduced to real wages. The indexes of both nominal and real wages

TABLE IX

INDEXES OF NOMINAL AND REAL WAGES, JULY, 1870, 1880, AND 1891

| SOURCE OF DATA | PRIMARY INDEXES | | | DERIVED INDEXES | | |
|---|---|---|---|---|---|---|
| | 1870 | 1880 | 1891 | 1870 | 1880 | 1891 |
| Connecticut Mill | | | | | | |
| Nominal wages (average) | 1.375 | 1.207 | 1.50 | 100 | 88 | 109 |
| Prices | 1.47 | 1.09 | .82 | | | |
| Real wages | .935 | 1.11 | 1.83 | 100 | 118 | 196 |
| Mass. Cotton Mills | | | | | | |
| Nominal wages (base, 1860) | 166 | 154 | 172 | 100 | 93 | 104 |
| Prices | 140 | 105 | 79 | | | |
| Real wages | 118 | 147 | 217 | 100 | 125 | 184 |
| U. S.-Aldrich Report | | | | | | |
| Nominal wage (base, 1860) | 167.1 | 143.0 | 168.6 | 100 | 86 | 101 |
| Prices (base, 1860) | 144.4 | 104.9 | 94.4 | | | |
| Real wages | 116 | 136 | 179 | 100 | 118 | 154 |
| U. S. Bureau of Lab. Stat. | | | | | | |
| Nominal wage (base, 1913) | 67 | 60 | 69 | 100 | 90 | 103 |
| Prices (base, 1913) | 147 | 109 | 82 | | | |
| Real wages | 46 | 55 | 84 | 100 | 121 | 185 |

are next reduced to a base of 1870, in which form they may be readily compared. It will be seen that the indexes of real wages in the Connecticut mill, based on very slender data though they are, do not differ markedly from the others.

A more complete statement of the results of Professor Howard's study, and of the Bureau of Labor Statistics' index, is presented in Table X.  As in the

TABLE X

INDEXES OF WAGES AND PRICES, 1870-1920

| Year | MASSACHUSETTS | | | UNITED STATES | | |
|------|------------------|----------------------|----------------|----------------------|----------------------|----------------|
|      | Wages (Weekly) | Wholesale Prices | Real Wages | Wages (Hr. Rates) | Wholesale Prices | Real Wages |
| 1870 | 166 | 140 | 50 | 67 | 147 | 46 |
| 1871 | 177 | 130 | 58 | 68 | 136 | 50 |
| 1872 | 183 | 135 | 58 | 69 | 141 | 49 |
| 1873 | 178 | 134 | 56 | 69 | 140 | 49 |
| 1874 | 163 | 128 | 54 | 67 | 133 | 50 |
| 1875 | 150 | 120 | 53 | 67 | 125 | 54 |
| 1876 | 145 | 111 | 55 | 64 | 116 | 55 |
| 1877 | 142 | 108 | 55 | 61 | 113 | 54 |
| 1878 | 145 | 98 | 63 | 60 | 102 | 59 |
| 1879 | 144 | 94 | 65 | 59 | 98 | 60 |
| 1880 | 154 | 105 | 62 | 60 | 109 | 55 |
| 1881 | 149 | 102 | 62 | 62 | 106 | 58 |
| 1882 | 157 | 103 | 64 | 63 | 108 | 58 |
| 1883 | 158 | 98 | 69 | 64 | 102 | 63 |
| 1884 | 155 | 89 | 74 | 64 | 92 | 70 |
| 1885 | 150 | 82 | 77 | 64 | 86 | 74 |
| 1886 | 153 | 80 | 80 | 64 | 84 | 76 |
| 1887 | 160 | 81 | 83 | 67 | 84 | 80 |
| 1888 | 164 | 84 | 83 | 67 | 87 | 77 |
| 1889 | 169 | 81 | 89 | 68 | 84 | 81 |
| 1890 | 173 | 80 | 91 | 69 | 81 | 85 |
| 1891 | 172 | 79 | 92 | 69 | 82 | 84 |
| 1892 | 172 | 75 | 97 | 69 | 76 | 91 |
| 1893 | 180 | 75 | 102 | 69 | 77 | 90 |
| 1894 | 168 | 68 | 104 | 67 | 69 | 97 |
| 1895 | 165 | 66 | 105 | 68 | 70 | 97 |
| 1896 | 175 | 64 | 116 | 69 | 66 | 105 |
| 1897 | 174 | 64 | 116 | 69 | 67 | 103 |
| 1898 | 171 | 66 | 110 | 69 | 69 | 100 |
| 1899 | 164 | 72 | 97 | 70 | 74 | 95 |
| 1900 | 189 | 78 | 102 | 73 | 80 | 91 |
| 1901 | 190 | 77 | 104 | 74 | 79 | 94 |
| 1902 | 191 | 80 | 101 | 77 | 85 | 91 |
| 1903 | 197 | 81 | 103 | 80 | 85 | 94 |
| 1904 | 196 | 80 | 103 | 80 | 86 | 93 |
| 1905 | 200 | 82 | 103 | 82 | 85 | 96 |
| 1906 | 216 | 87 | 105 | 85 | 88 | 97 |
| 1907 | 240 | 92 | 111 | 89 | 94 | 95 |
| 1908 | 228 | 87 | 111 | 89 | 91 | 98 |
| 1909 | 211 | 90 | 99 | 90 | 97 | 93 |
| 1910 | 209 | 93 | 95 | 93 | 99 | 94 |
| 1911 | 207 | 92 | 96 | 95 | 95 | 100 |
| 1912 | 223 | 95 | 99 | 97 | 101 | 96 |
| 1913 | 227 | 96 | 100 | 100 | 100 | 100 |
| 1914 | 229 | 95 | 102 | 102 | 100 | 102 |
| 1915 | (Base | (Base | | 103 | 101 | |
| 1916 | 1860) | 1860) | | 111 | 124 | |
| 1917 | | | | 128 | 176 | |
| 1918 | | | | 162 | 196 | |
| 1919 | | | | 184 (Spring) | 212 | |
| 1920 | | | | 234 (Summer) | 243 | |

preceding table, an index of real wages is derived by the use of an index of wholesale prices. The price index shown for Massachusetts is merely an adaptation of the Bureau of Labor Statistics' data for the United States. The index of real wages for Massachusetts has been changed from a base of 1860, as first derived, to a base of 1913, in order to allow of comparisons with the corresponding index for the United States. Both indexes point to the fact that real wages have about doubled in the interval from 1870 to 1914, but that the greater part of this increase came before 1890.

**Wholesale and Retail Prices.** A question may be raised regarding the validity of using an index of wholesale prices as a measure of changes in the cost of living. Unfortunately, no adequate index of retail prices covering the years here studied is available. An index of wholesale prices has therefore been substituted. It is a well established fact, however, that wholesale prices swing in the same direction and at nearly the same time as retail prices, but that they move somewhat more extremely. In the course of the usual moderate cyclic changes, therefore, the former will parallel the latter closely enough to serve as a substitute. But when the price swings are extremely low, the substitution will doubtless give an exaggerated rise in real wages. Such is evidently the case in the decade from 1890 to 1900, when prices fell to the lowest point in the century. The apparent rise in real wages at that time should therefore be discounted, though just how much, cannot be accurately determined. The opposite result is obtained during the great upswing of prices from 1914 to 1920. For these

years, however, adequate data on the cost of living are obtainable. In Table XI estimates derived from such data are used in the place of wholesale prices, and real wages are then computed.[1]

TABLE XI

INDEXES OF WAGES AND COST OF LIVING
UNITED STATES, 1913-1920

(Estimated from Bureau of Labor Statistics' data)

| Year | Wages | Cost of Living | Real Wages |
|------|-------|----------------|------------|
| 1913 | 100 | 100 | 100 |
| 1914 | 102 | 100 | 102 |
| 1915 | 103 | 100 | 103 |
| 1916 | 111 | 110 | 101 |
| 1917 | 128 | 134 | 95 |
| 1918 | 162 | 154 | 105 |
| 1919 | 200 | 180 | 111 |
| 1920 | 225 | 211 | 107 |

**Indexes of the Cost of Living.** The computation of an index of the cost of living involves many difficulties, both theoretical and practical. To begin with, the units employed often vary in quality, and are difficult to standardize. Hence the first step in the computation is the drawing up of a selected list of articles of staple grades. These articles must be sufficient in number

[1] It is not intended that the wage data here studied shall be taken as a final measurement of the course of real wages. They were compiled chiefly with the purpose in mind of illustrating certain methods of work. But it is the opinion of the writer, based on a study of such material as is available, that they represent a passably good estimate. Other studies, however, have purported to show a decline in real wages since the last decade of the nineteenth century. An interesting study of this sort appears in the *American Economic Review,* September, 1921. In this study the cost of living is assumed to be measured by retail food prices. But this is a very questionable measure, inasmuch as retail food prices have risen about as rapidly as wholesale prices since the period of agricultural depression. This abnormal rise, of course, makes real wages by comparison appear to fall. The study also uses an index of hour rates which purports to show a slower rise than is shown by the index published by the Bureau of Labor Statistics. It should be noted that the data here discussed do not cover wages of government employes.

and importance to represent fairly all the necessities commonly purchased by the average working-class family. After this has been done, the prices of these articles as sold in representative stores must be tabulated. If the index is to cover a considerable territory, data must be gathered from a number of localities. The common average of the prices of each article at a given date is found. Under certain conditions, as when the extreme items are likely to be in error, the median is preferable in place of the common average. Since the work of collecting and tabulating prices requires elaborate organization, it is done on a large scale by only a few agencies. One of these is the National Industrial Conference Board.[1] Another is a commission established by the Legislature of the State

[1] The National Industrial Conference Board is affiliated with the National Association of Manufacturers and other similar organizations, and has its headquarters in New York. It publishes among other things an excellent index of increases in the cost of living in the United States. This index is issued promptly each month, and is the best available source by which current changes may be measured. By permission, the indexes for the year 1920 are here reprinted. The base is July, 1914, and the figures show the per cent rise.

| Month, 1920 | Cost of Living | Food | Shelter | Clothing | Fuel and Light | Sundries |
|---|---|---|---|---|---|---|
| January | 90.2 | 97 | 43 | 170 | 49 | 77 |
| February | 93.5 | 101 | 45 | 177 | 49 | 78 |
| March | 94.8 | 100 | 49 | 177 | 49 | 83 |
| April | 96.6 | 100 | 50 | 188 | 51 | 83 |
| May | 101.6 | 111 | 51 | 187 | 55 | 83 |
| June | 103.0 | 115 | 51 | 176 | 61 | 85 |
| July | 104.5 | 119 | 58 | 166 | 66 | 85 |
| August | 103.2 | 119 | 58 | 155 | 69 | 85 |
| September | 99.4 | 107 | 59 | 155 | 78 | 83 |
| October | 97.3 | 103 | 59 | 148 | 83 | 90 |
| November | 93.1 | 93 | 66 | 128 | 100 | 92 |
| December | 90.0 | 93 | 66 | 105 | 100 | 92 |

Of these indexes the only one which shows a further rise up to October, 1921, is that for shelter. This registers 71% during March to June, inclusive, and 69% during the four months following. Fuel and Light and Sundries begin to decline after January. Food shows an increase for four months following a minimum of 45% in June, but this change is reflected in only a minor degree in the aggregate index. The apex of the post-war boom is, then, according to this index, in July, 1920.

TABLE XII

FOOD COST OF LIVING PER FAMILY, JUNE, 1920, COMPARED WITH AVERAGE COST IN 1913

| ARTICLE OF FOOD | (1) CONSUMPTION | (2) PRICE PER UNIT 1913 | (3) COST 1913 | (4) PRICE PER UNIT JUNE 1920 | (5) COST JUNE 1920 | (6) RELATIVE PRICE | (7) "WEIGHT" | (8) EXTENSION |
|---|---|---|---|---|---|---|---|---|
| Sirloin steak | 32 lbs. | $ .254 | $8.128 | $ .461 | $14.752 | 182% | 38 | 6916% |
| Round steak | 32 lbs. | .223 | 7.136 | .426 | 13.632 | 191 | 33 | 6303 |
| Rib roast | 31 lbs. | .198 | 6.138 | .348 | 10.788 | 176 | 28 | 4928 |
| Chuck roast | 31 lbs. | .160 | 4.960 | .278 | 8.618 | 174 | 23 | 4002 |
| Plate beef | 23 lbs. | .121 | 2.783 | .190 | 4.370 | 157 | 13 | 2041 |
| Pork chops | 36 lbs. | .210 | 7.560 | .408 | 14.688 | 194 | 35 | 6790 |
| Bacon | 17 lbs. | .270 | 4.590 | .539 | 9.163 | 200 | 21 | 4200 |
| Ham | 22 lbs. | .269 | 5.918 | .577 | 12.694 | 215 | 27 | 5805 |
| Lard | 34 lbs. | .158 | 5.372 | .293 | 9.962 | 185 | 25 | 4625 |
| Hens | 23 lbs. | .213 | 4.899 | .460 | 10.580 | 216 | 23 | 4968 |
| Eggs | 61 doz. | .345 | 21.045 | .536 | 32.696 | 155 | 98 | 15190 |
| Butter | 66 lbs. | .383 | 25.278 | .672 | 44.352 | 175 | 117 | 20475 |
| Cheese | 12 lbs. | .221 | 2.652 | .418 | 5.016 | 189 | 12 | 2268 |
| Milk | 337 qts. | .089 | 29.993 | .162 | 54.594 | 182 | 139 | 25298 |
| Bread | 531 lbs. | .056 | 29.736 | .118 | 62.658 | 211 | 138 | 29118 |
| Flour | 264 lbs. | .033 | 8.712 | .088 | 23.232 | 267 | 40 | 10680 |
| Corn meal | 54 lbs. | .030 | 1.620 | .069 | 3.726 | 230 | 8 | 1840 |
| Rice | 35 lbs. | .087 | 3.045 | .187 | 6.545 | 215 | 14 | 3010 |
| Potatoes | 704 lbs. | .017 | 11.968 | .103 | 72.512 | 606 | 55 | 33330 |
| Sugar | 147 lbs. | .055 | 8.085 | .267 | 39.249 | 485 | 38 | 18430 |
| Coffee | 40 lbs. | .298 | 11.920 | .492 | 19.680 | 165 | 55 | 9075 |
| Tea | 8 lbs. | .544 | 4.352 | .741 | 5.928 | 136 | 20 | 2720 |
| Total | | | $215.890 | | ) $479.435 | | 1000 | ) 222012% |
| | | | | | 222% | | | 222% |

of Massachusetts. But doubtless the best known and most authoritative work is done by the Bureau of Labor Statistics. The methods used in combining average prices into a single index may be illustrated by the Bureau of Labor Statistics' index of the food cost of living.

**An Index of Food Prices.** In computing an index of the food cost of living, the Bureau of Labor Statistics has listed 22 important articles of food, as shown in Table XII. The use of this list was begun in January, 1913, and was continued to January, 1921, when a revised list of 43 articles was substituted. Prices on these articles were compiled monthly from a number of important cities selected from the various territorial divisions of the United States. The averages of these prices for the year 1913 and for the month of June, 1920, are shown in Table XII, together with the annual consumption per workingman's family as ascertained by a special investigation, which is here assumed to apply directly to 1913. From these data the increase in the food cost of living from the pre-war level to the climax of prices in 1920 may be computed.

**The Aggregate Method.** The two standard methods of making the computation are shown in Table XII. The first and simplest of these is known as the aggregate method, and is illustrated in the first five columns of the table. The process consists merely in finding the total cost of a year's supply of the given articles at the two contrasted dates (columns 3 and 5.) The total cost in June, 1920 ($479.44), is then divided by the total cost in 1913 ($215.89). The result shows that the former cost was 222% as compared with 100% in the base year, 1913—an increase of 122%. While the

totals do not actually represent more than about two-thirds of the average annual food cost per family, yet they doubtless are representative enough to give a close approximation to the correct percentage increase.

**The Proportional Expenditure Method.**  A second and more complex process, known as the proportional expenditure method, is illustrated in columns 6, 7, and 8.  By this method the relative price in June, 1920, as compared with 1913 is found (column 6).  By so doing, all prices, whether large or small, are obviously put on the same basis, since each is based on 100% in 1913.  These relative prices are next averaged by a process of weighting similar to that explained in Chapter II in connection with Table IV.  The weights, as shown in column 7, measure the relative importance in a work-ingman's budget of an annual supply of each article.  To illustrate, the first weight is obtained by dividing $8.128 by $215.89.  This gives approximately 3.8%, but both the decimal point and the per cent sign are dropped as being unnecessary in a ratio.  In the same way the second weight, 33, is obtained by dividing $7.136 by $215.89, and so on for the remainder of the weights.  The accuracy of the work may be checked by means of the total, which should come to approximately 1000.  In the computation of such weights, it is usually unnecessary to strive for extreme accuracy; as, for example, by carrying the figures to several decimal places.  A considerable margin of error may be present in the weights without materially affecting the weighted average.

**A Comparison of the Two Methods.**  It will be observed that the result by the proportional expenditure method is exactly the same in this instance as that

obtained by the aggregate method. But this would not
always be the case. Where there is a difference, pref-
erence is likely to be given to the result obtained by the
former method. The advantage claimed for this
method arises from the fact that it is not practicable
to make frequent revisions of the consumption esti-
mates. When an estimate has become somewhat out
of date, its inaccuracies may have considerable effect
upon the results obtained by the aggregate method.
But it is assumed that the proportional expenditure
weights, based as they are upon both the prices and
the consumption of a specific period, will remain rela-
tively accurate for a longer period than will the con-
sumption estimates taken alone. Hence their use is
thought to give somewhat more dependable results.
In practice it will be found that the proportional ex-
penditure method is not as tedious as it seems in the
illustration. The weights, when once· obtained, may
be used unchanged for a long period; and the finding
of the relative prices does not usually mean additional
work, since they are in any case desirable for purposes
of comparison.

The formula for the price index by the aggregate
method is:

$$P_n = \frac{\Sigma p_n q_x}{\Sigma p_0 q_x}$$

in which

$P_n$ = price index for the relative period

$p_n$ = prices for the relative period

$q_x$ = quantities consumed during a given period,
preferably the base period

$p_0$ = prices for the base period

The process may be described briefly as a comparison

of two price averages, both obtained by weighting for quantities.

The formula for a price index by the proportional expenditure method is:[1]

$$P_n = \frac{\Sigma \frac{p_n}{p_o} \times p_x q_x}{\Sigma p_x q_x}$$

This process may be described as an average of relative prices weighted for the value consumed during some specified period, often prior to the beginning of the interval over which the price comparison is being made. In connection with both formulas, it must be understood that the factors are taken distributively; that is, only prices and weights belonging to the same article are multiplied.[2]

An inspection of the formulas will show why the two results obtained in Table XII are alike. As is often the case, the year indicated by the subscript $x$, to which the quantities and weights are assumed to belong, is identical with the base year indicated by the

[1] The formula as thus stated does not indicate the reduction of the proportional expenditure weights to percentages. The value is, of course, unaffected by such reduction; and for purposes of later comparison the form as given is preferable.

[2] Use is sometimes made of the harmonic mean in finding the average of the relatives. This average is found by taking the reciprocals of the relatives, computing their average, and then taking the reciprocal of this average. It may easily be seen that the effect of taking the reciprocals of the relatives is to reverse the base; that is, the later year becomes the base instead of the earlier year. If weights are used it is therefore preferable that they be derived from the data of the later year in finding the harmonic mean. Taking the reciprocal of the average again reverses the bases. But the index obtained by the usual direct method will not be quite the same as that obtained by the harmonic mean. This will ordinarily be the case whether the weights are shifted to the later base, or not; or indeed whether any weights are employed or not. The same distinction between the arithmetic and the harmonic mean is well brought out by taking a simple average of a given set of prices, and comparing it with the harmonic mean of the same prices. The latter is the reciprocal of the average of the quantities of each commodity purchasable for one dollar.

TABLE XII-A

ANNUAL CONSUMPTION PER FAMILY, AND AVERAGE RETAIL PRICES OF SPECIFIED ARTICLES OF FOOD, UNITED STATES, 1913-1920, AND JUNE, 1921.

| ARTICLE OF FOOD | CONSUMPTION 1901 | AVERAGE RETAIL PRICES | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 1913 | 1914 | 1915 | 1916 | 1917 | 1918 | 1919 | 1920 | JUNE 1921 |
| Sirloin steak | 70 lbs. | $.254 | $.259 | $.257 | $.273 | $.315 | $.389 | $.417 | $.347 | $.400 |
| Round steak | 70 lbs. | .223 | .236 | .230 | .245 | .290 | .369 | .389 | .385 | .356 |
| Rib roast | 70 lbs. | .198 | .204 | .201 | .212 | .249 | .307 | .325 | .332 | .298 |
| Chuck roast | 70 lbs. | .160 | .167 | .161 | .171 | .209 | .266 | .270 | .262 | .216 |
| Plate beef | 70 lbs. | .121 | .126 | .121 | .128 | .157 | .206 | .202 | .183 | .141 |
| Pork chops | 114 lbs. | .210 | .220 | .203 | .227 | .319 | .390 | .423 | .423 | .341 |
| Bacon | 55 lbs. | .270 | .275 | .269 | .287 | .410 | .529 | .554 | .523 | .429 |
| Ham | 55 lbs. | .269 | .273 | .261 | .294 | .382 | .479 | .534 | .555 | .459 |
| Lard | 84 lbs. | .158 | .156 | .148 | .175 | .276 | .333 | .369 | .295 | .162 |
| Hens | 68 lbs. | .213 | .218 | .208 | .236 | .286 | .377 | .411 | .447 | .386 |
| Eggs | 85 doz. | .345 | .353 | .341 | .375 | .481 | .569 | .628 | .681 | .350 |
| Butter | 117 lbs. | .383 | .362 | .358 | .394 | .487 | .577 | .678 | .701 | .402 |
| Cheese | 16 lbs. | .221 | .229 | .233 | .258 | .332 | .359 | .426 | .416 | .295 |
| Milk | 355 qts. | .089 | .089 | .088 | .091 | .112 | .139 | .155 | .167 | .142 |
| Bread | 225 lbs. | .056 | .063 | .070 | .073 | .092 | .098 | .100 | .115 | .098 |
| Flour | 454 lbs. | .033 | .034 | .042 | .044 | .070 | .067 | .072 | .081 | .059 |
| Corn meal | 227 lbs. | .030 | .032 | .033 | .034 | .058 | .068 | .064 | .065 | .045 |
| Rice | 25 lbs. | .087 | .088 | .091 | .091 | .104 | .129 | .151 | .174 | .088 |
| Potatoes | 882 lbs. | .017 | .018 | .015 | .027 | .043 | .032 | .038 | .063 | .027 |
| Sugar | 269 lbs. | .055 | .059 | .066 | .080 | .093 | .097 | .113 | .194 | .078 |
| Coffee | 47 lbs. | .298 | .297 | .300 | .299 | .302 | .305 | .433 | .470 | .357 |
| Tea | 11 lbs. | .544 | .546 | .545 | .546 | .582 | .648 | .701 | .733 | .683 |

subscript *o*. The numerator of the second formula therefore reduces to the same form as that of the first formula, and the denominators also have the same value. But the assumption that the quantities as given apply directly to the year 1913 was adopted merely to simplify the table for purposes of illustration. As a matter of fact, these quantities were obtained by an investigation made in 1918, and were not put into use until the beginning of 1921. The consumption estimates actually used by the Bureau of Labor Statistics from 1913 to 1920, inclusive, are shown in Table 12-A, together with the annual retail food prices for the same period, and for June, 1921. It will be seen that these estimates go back to the year 1901.

**Limitations.** Some of the limitations of the foregoing methods of computing changes in the cost of living may be easily seen. It is evident that the assuming of an unvarying consumption may sometimes result in material inaccuracies. If there is a very uneven rise in prices, buyers will begin to substitute the cheaper article for the more expensive, wherever this can be properly done. The effect on the one hand will be to moderate the unevenness of price changes, but on the other hand it will make material alterations in the budget. But the frequent revision of consumption data, and the complexity of methods of computation, would involve more labor than at present seems practicable to allow. Besides, the revision of consumption data raises another problem. Changing quantities may in part reflect rising standards of comfort whereas the term, ''cost of living,'' implies provision only for those necessities which are required to maintain the social efficiency of the family. Strictly

speaking, the measurement of changes in such costs would call for the dietician's units of protein, carbohydrates, fats, and calories, and would require very detailed analyses. Hence, as far as the cost of living is concerned, we are driven back to the comparatively simple methods actually in use. The results obtained by such methods must, of course, always be taken as approximations only.

**Combining the Partial Indexes.** In measuring changes in the entire cost of living, the Bureau of Labor Statistics makes an index for several groups of commodities, in addition to food. The process of combining these partial indexes into a single measure is an application of the proportional expenditure method. The budgetary studies of 1918 furnish the weights now in use. The partial indexes in June, 1920, the weights, and the process of finding a single index, may be seen in the following table:

| Item of Expenditure | Index June, 1920 (base, 1913) | Weight (per cent of budget) | Extension |
|---|---|---|---|
| Food | 219.0 | 38.2 | 8365.80 |
| Clothing | 287.5 | 16.6 | 4772.50 |
| Housing | 134.9 | 13.4 | 1807.66 |
| Fuel and light | 171.9 | 5.3 | 911.07 |
| Furniture and furnishings | 292.7 | 5.1 | 1492.77 |
| Miscellaneous | 201.4 | 21.3 | 4289.82 |
| | | 99.9 ) | 21639.62 |
| | | | 216.5 |

**Indexes of Wholesale Prices.** Though neither the aggregate nor the proportional expenditure method is theoretically perfect, yet one or the other is used in practically all the price indexes in common use. In

adapting the latter method to wholesale prices, the weights are usually made to reflect the proportional value produced, rather than family consumption, but the principle is essentially the same. Applying this method, the Bureau of Labor Statistics compiles an elaborate monthly index of wholesale prices. The number of items tabulated, the base used, and other details of the work have been varied from time to time, but the partial results have been combined into a single index having the same base. For several years prior to the war, the base used was an average of the decade 1890-99, which was a period of unusually low prices. But in more recent years, 1913 has been employed because it may be considered representative of normal conditions immediately preceding the war. As compiled for January, 1920, the index is derived from a weighted average of 327 price quotations. The items are classified under nine headings, each having its own partial index. [1] This is the most com-

[1] Annual group index numbers for the years 1890-1920 will be found in the *Monthly Labor Review,* Feb., 1921, page 45. The figures for recent dates are as follows:

GROUP INDEX NUMBERS—UNITED STATES—BUREAU OF LABOR STATISTICS

For the Years 1913-1920, and Sept., 1920 and 1921

| Period | Farm products | Food, etc. | Cloths and clothing | Fuel and lighting | Metals and metal products | Lumber and building material | Chemicals and drugs | House-furnishing goods | Miscel-laneous |
|---|---|---|---|---|---|---|---|---|---|
| 1913 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 1914 | 103 | 103 | 98 | 96 | 87 | 97 | 101 | 99 | 99 |
| 1915 | 105 | 104 | 100 | 93 | 97 | 94 | 114 | 99 | 99 |
| 1916 | 122 | 126 | 128 | 119 | 148 | 101 | 159 | 115 | 120 |
| 1917 | 189 | 176 | 181 | 175 | 208 | 124 | 198 | 144 | 155 |
| 1918 | 220 | 189 | 239 | 163 | 181 | 151 | 221 | 196 | 193 |
| 1919 | 234 | 210 | 261 | 173 | 161 | 192 | 179 | 236 | 217 |
| 1920 | 218 | 239 | 302 | 238 | 186 | 308 | 210 | 366 | 236 |
| Sept. 1920 | 210 | 223 | 278 | 284 | 192 | 318 | 222 | 371 | 239 |
| Sept. 1921 | 122 | 146 | 187 | 178 | 120 | 193 | 162 | 223 | 146 |

prehensive study of wholesale prices published, but for business uses it has the disadvantage of appearing two or three months late.

One of the most widely used commercial indexes of wholesale prices is Bradstreet's. This index is based on about one hundred important articles, for which prices are easily available in the principal business centers. It is published promptly each month, and is therefore valuable to the business man who desires to keep in touch with the immediate trend of the market. The aggregate method of computation is employed, and the result is given merely as a sum of money, which may be reduced to any desired base. The lack of a definite system of weights is a defect of this index, but it is partially compensated for by a careful selection of items, and a repetition of important articles by a quotation for more than one grade.

Dun's monthly index of wholesale prices is perhaps not as widely known as Bradstreet's, but it appears to be more scientifically compiled. It is based upon approximately 300 price quotations, which are grouped under seven heads, and finally combined into a single index, the result being stated in dollars. Prices are weighted in accordance with estimated per capita consumption. Just how the weights are applied is not apparent. Commercial houses, as a rule, do not publish the details of their statistical work.

The Federal Reserve Board has begun the publication of a monthly index of wholesale prices, based on about 90 commodities. It is intended for use particularly in international comparisons, but its value is not limited to this field.[1] Still another index intended

[1] The Federal Reserve Board classifies its data so as to present indexes

to measure wholesale price movements is published by the Babson Statistical Organization. This index is compiled monthly from quotations on ten basic commodities, and in spite of its narrow base, serves a useful purpose. A number of other less complete indexes might be mentioned, some of which appear weekly. One of these is the *Annalist's* weekly index of wholesale food, a weighted average of 25 prices. It is obvious that there may be many price indexes giving somewhat different results, yet each one valid in its own sphere. The purpose a given index is to serve will always control the selection of items and the weighting.

As a means of comparing some of the indexes just mentioned, Table XIII is given, showing several wholesale price indexes for the year 1913, and by months for 1920. The data for 1920 are of interest because they contain the peak of prices for the war and postwar cycle.[1]

TABLE XIII

INDEXES OF WHOLESALE PRICES

| Period | Bureau of Labor Statistics | Bradstreet's | Dun's | Federal Reserve Board | Babson's |
|--------|-----|-----|-----|-----|-----|
| 1913 | 100 | $9.2115 | $120.8865 | 100 | $1.26 |
| 1920 | | | | | |
| Jan. | 248 | 20.3638 | 247.394 | 242 | 3.30 |
| Feb. | 249 | 20.8690 | 253.748 | 242 | 3.44 |
| Mar. | 253 | 20.7950 | 253.016 | 248 | 3.59 |
| Apr. | 265 | 20.7124 | 257.901 | 263 | 3.61 |
| May | 272 | 20.7341 | 263.332 | 264 | 3.66 |
| June | 269 | 19.8752 | 262.149 | 258 | 3.71 |
| July | 262 | 19.3528 | 260.414 | 250 | 3.60 |
| Aug. | 250 | 18.8273 | 252.288 | 234 | 3.43 |
| Sept. | 242 | 17.9746 | 248.257 | 226 | 3.39 |
| Oct. | 225 | 16.9094 | 237.341 | 208 | 3.25 |
| Nov. | 207 | 15.6750 | 227.188 | 190 | 2.98 |
| Dec. | 189 | 13.6263 | 211.628 | 171 | 2.75 |

for goods imported and exported, and producers' and consumers' goods, as well as for certain commodity groups. These indexes are published monthly in the *Federal Reserve Bulletin*.

[1] In comparing these data, it should be noted that the commercial indexes (Bradstreet's, Dun's, and Babson's) are based on price quotations for the first of each month.

**Other Indexes.** In addition to indexes of commodity prices, many other indexes, measuring various phases of business activity, are published. Speculative and investment activities are measured by indexes of stock and bond prices. A good index of stock prices is difficult to make because of the continually changing status of the corporations issuing the stocks. The indexes now in common use are based on quotations of standard shares listed on the New York Stock Exchange. Railroad and industrial shares are usually compiled separately, and the two sets combined into a composite index. Financial and productive activities in the general markets are measured by a variety of data, such as the number of shares traded on the New York Stock Exchange, bank clearings in New York and in the country as a whole, bank deposits, money in circulation, gold movements, the interest rate, the production of pig iron, the number of building permits issued in leading cities, the number and extent of business failures, and the balance of foreign trade. Labor conditions are measured by wage and employment data, examples of which are the reports of the New York State Industrial Commission. Retail trade is indicated by reports of department store sales. Most of these various indexes as they appear in the financial papers, consist of mere statements of periodical figures. Certain phases of their statistical elaboration will be considered later under the subject of trends.

### *REFERENCES*

Barnett, George E., ''Index Numbers of the Total Cost of Living,'' *Quarterly Journal of Economics,* February, 1921, pp. 240-263.

Bowley, Arthur L., "The Measurement of Changes in the Cost of Living" (and discussion), *Journal of the Royal Statistical Society*, May, 1919, pp. 343-372.

Fisher, Irving, *Stabilizing the Dollar*, Chapters I-III.

Howard, Stanley E., *The Movement of Wages in the Cotton Manufacturing Industry of New England.*

Meeker, Royal, "Some Features of the Statistical Work of the Bureau of Labor Statistics," *Quarterly Publications of the American Statistical Association*, March, 1915, pp. 431-441.

Mitchell, Wesley C., *Index Numbers of Wholesale Prices in the United States and Foreign Countries*, Bulletin No. 173 (Whole Number), Bureau of Labor Statistics, July, 1915.

Secrist, Horace, *An Introduction to Statistical Methods,* Chapters IX and X.

Secrist, Horace, *Readings and Problems in Statistical Methods,* Chapter VIII.

Stewart, Walter W., "Prices During the War," *Quarterly Publications of the American Statistical Association*, September, 1920, pp. 305-313.

## EXERCISES

1. Reduce the three indexes shown on page 53 to a base consisting of the average of each series, respectively.

2. Reduce Bradstreet's, Dun's and Babson's indexes of wholesale prices for 1920 (page 65) to a 1913 base. Compare the divergence of these indexes, and the others given in the same table, for the month of May by the use of the coefficient of average deviation.

3. (a) Plot on the same sheet of semi-logarithmic paper the two indexes of real wages given in Table X, page 51.
   (b) Plot together on the same chart the indexes of wages and cost of living shown in Table XI, page 53.

4. Assuming that the ratio of American to European commodity prices—both price levels being stated in indexes having 1913 as a base—measures the relative depreciation of European currencies, find the theoretical value of these currencies for September, 1921, using the data given below (*Federal Reserve Bulletin,* Nov., 1921).

Express the market prices of exchange as percentages of the theoretical values.

| Country | Index of Wholesale Prices September, 1921 | Exchange in New York. (Cables, per cent of par) |
|---|---|---|
| United States ...... | 143 | — |
| United Kingdom ... | 191 | 76.5 |
| France ........... | 344 | 37.7 |
| Italy ............. | 580 | 21.8 |
| Germany ......... | 1777 | 4.0 |
| Sweden ........... | 182 | 81.3 |
| Norway ........... | 287 | 48.0 |
| Denmark .......... | 224 | 65.9 |

5. The following table gives a monthly index of wholesale prices obtaining in the United Kingdom ("Statist" index) and the average monthly price in New York of sterling exchange (cables), for the year 1920. Using the Federal Reserve Board index (page 65) as a measure of the price level in the United States, find the theoretical value of the pound sterling (par, $4.8665) each month in terms of American money. Compare the results thus obtained with the cost of sterling exchange by graphing both series. (The graph should be drawn so as to show the zero line at the base. Why?)

| 1920 | Statist index (Base, 1913) | Sterling cables (New York) |
|---|---|---|
| January ................ | 288 | $3.68 |
| February ............... | 306 | 3.39 |
| March ................. | 307 | 3.72 |
| April .................. | 313 | 3.93 |
| May ................... | 305 | 3.85 |
| June .................. | 300 | 3.95 |
| July .................. | 299 | 3.86 |
| August ................ | 298 | 3.63 |
| September ............. | 292 | 3.52 |
| October ............... | 282 | 3.47 |
| November ............. | 263 | 3.43 |
| December ....... ....... | 243 | 3.63 |

6. From the data of Table XII-A, page 60, find index numbers of the food cost of living, 1913-1920, inclusive, and June, 1921; base, 1913. Compare the results with the following index published by the Bureau of Labor Statistics:

| Year | Index of food costs |
|------|---------------------|
| 1913 | 100 |
| 1914 | 102 |
| 1915 | 101 |
| 1916 | 114 |
| 1917 | 146 |
| 1918 | 168 |
| 1919 | 186 |
| 1920 | 203 |
| June, 1921 | 144* |

* Based on 43 articles.

7. From the average prices given below, find the relative prices in 1918 as compared with 1913. Find also the weighted average of these "relatives," applying the weights given.

|  | 1913 | 1918 | Weights |
|--|------|------|---------|
| Wheat, bushel | $1.04 | $2.31 | 4 |
| Corn, bushel | .71 | 1.84 | 10 |
| Cotton, pound | .13 | .32 | 5 |
| Iron, ton | 14.90 | 36.52 | 3 |
| Copper, pound | .16 | .25 | 1 |

8. On the basis of the following prices of five articles of food, and the relative importance of each in the family budget, find the increase in the food cost of living between the dates given:

| Article | Price in 1913 | Price in 1920 | Importance |
|---------|---------------|---------------|------------|
| Steak, pound | $0.22 | $0.40 | 70 |
| Milk, quart | .09 | .17 | 140 |
| Bread, pound | .06 | .12 | 140 |
| Butter, pound | .38 | .70 | 115 |
| Sugar, pound | .06 | .20 | 40 |

9. Compute a set of proportional expenditure weights on the basis of 1901 consumption and 1913 prices. Apply these weights to the relatives given in column 6, Table XII, page 55.

10. From the following table, compute an index of agricultural real wages having 1913 as its base. Compare the results with the corresponding index of hour rates in the United States by graphing both on the same chart for the years in which farm wages are given. (The base line of the graph should show each year from 1875 to 1920. Why?)

WAGES OF CERTAIN CLASSES OF MALE FARM LABOR BY THE
MONTH WITHOUT BOARD

("Monthly Labor Review," July, 1920, and March, 1921.)

| Year | Wage |
|------|------|
| 1875 | 19.87 |
| 1879 | 16.42 |
| 1882 | 18.94 |
| 1885 | 17.97 |
| 1888 | 18.24 |
| 1890 | 18.33 |
| 1892 | 18.60 |
| 1893 | 19.10 |
| 1894 | 17.74 |
| 1895 | 17.69 |
| 1898 | 19.38 |
| 1899 | 20.23 |
| 1902 | 22.14 |
| 1910 | 27.50 |
| 1911 | 28.77 |
| 1912 | 29.58 |
| 1913 | 30.31 |
| 1914 | 29.88 |
| 1915 | 30.15 |
| 1916 | 32.83 |
| 1917 | 40.43 |
| 1918 | 48.80 |
| 1919 | 56.29 |
| 1920 | 64.95 |

11. The Bureau of Labor Statistics found the increases in the cost of living for various classes of commodities in the United States, from 1913 to the year and month indicated to have been as follows:

| Item of Expenditure | Per Cent of Increase | | | | | | |
| | Dec. 1914 | Dec. 1915 | Dec. 1916 | Dec. 1917 | Dec. 1918 | Dec. 1919 | Dec. 1920 |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Food ............... | 5.0 | 5.0 | 26.0 | 57.0 | 87.0 | 97.0 | 78.0 |
| Clothing ........... | 1.0 | 4.7 | 20.0 | 49.1 | 105.3 | 168.7 | 158.5 |
| Housing ........... | 0.0 | 1.5 | 2.3 | .1 | 9.2 | 25.3 | 51.1 |
| Fuel and light....... | 1.0 | 1.0 | 8.4 | 24.1 | 47.9 | 56.8 | 94.9 |
| Furniture and furnishings........ | 4.0 | 10.6 | 27.8 | 50.6 | 113.6 | 163.5 | 185.4 |
| Miscellaneous ....... | 3.0 | 7.4 | 13.3 | 40.5 | 65.8 | 90.2 | 108.2 |

Using the weights given on page 62, find the total increase in the cost of living, and the index of the same, for the periods named in the foregoing table. Compare the results with those given in the *Monthly Labor Review*, November, 1921, page 83.

12. Using price quotations obtained locally, and from catalogs from which local purchases are commonly made, find the increase in the cost of living between two given dates, preferably 1913 or 1914 and the present time. In finding the food index, make use of the proportional expenditure weights given in Table XII, page 55. For the other groups of items, make use of the following proportional expenditure weights quoted from Massachusetts House Report, No. 1500. Items for which quotations are not available may be dropped, or substitutions may be made.

### WEIGHTINGS IN THE CLOTHING INDEX

#### MEN'S

Overcoat ⎫
Suit       ⎬ ........................................ 39
Trousers ⎭
Shoes ........................................... 15
Hats ............................................ 4
Gloves .......................................... 6
Socks ........................................... 4
Shirts .......................................... 6
Collars ......................................... 2
Underwear ...................................... 6
Night Garments ................................. 2
                                               ——
Total........................................... 84

### WOMEN'S

| | |
|---|---|
| Suit <br> Topcoat <br> Street Dress | 27 |
| Underwear | 5 |
| Waists <br> Kimono <br> House Dress <br> Aprons <br> Night Gown <br> Underskirt | 18 |
| Shoes | 12 |
| Gloves | 3 |
| Hosiery | 2 |
| Corsets | 4 |
| Hats | 9 |
| Total | 80 |

### SHELTER INDEX

Obtain rentals of several representative homes for the two dates required, and find the average per cent increase.

### WEIGHTINGS IN THE FUEL INDEX

| | |
|---|---|
| Coal | 10 |
| Kerosene | 1 |
| Gas | 2 |
| Electricity | 2 |
| Total | 15 |

### WEIGHTINGS IN THE SUNDRY INDEX

| | |
|---|---|
| Ice | 15 |
| Carfare | 15 |
| Entertainment | 25 |
| Medicine | 25 |
| Insurance | 50 |
| Church | 30 |
| Tobacco, etc. | 20 |
| Reading | 10 |
| House furnishings | 45 |
| Organizations | 25 |
| Total | 260 |

Combine the various partial indexes by means of the following proportional expenditure weights:

Food ........................ 43.1
Shelter ...................... 17.7
Clothing ........... .......... 13.2
Fuel and light ................ 5.6
Sundries ..................... 20.4

# CHAPTER IV

## QUANTITY INDEXES AND THEIR USES

**Value and Quantity Indexes.** Students of general
market conditions are interested in changes in the
volume of production, since such changes have an inti-
mate relation to prices, wages, and business activity.
The attention of statisticians has therefore been re-
cently turned to the development of indexes of pro-
duction. Such indexes may be of two sorts: (1) an
index of value production, measuring the number of
dollars' worth of goods created in a given period of
time; and (2) an index of physical production, meas-
uring the same output primarily in such physical units
as pounds, bushels, and yards. Somewhat akin to the
latter is an index of the physical volume of trade,
which attempts to measure the total number of physical
units traded in a given period. Because of the sea-
sonal nature of a large part of industry, indexes of
production are generally based on the year as the unit
of time.[1]

Indexes of value production do not call for an ex-
tended discussion, since they are merely inventories of
representative annual output at the average current
prices. The process of finding them may be illus-

[1] One of the most practical uses of production index numbers, how-
ever, requires monthly or weekly data, corrected for seasonal variations.
These data are beginning to be used in connection with business
barometrics, as will be noted in the next chapter.

trated by multiplying each year's output of wheat, corn, cotton, pig iron, and copper, as shown in Table XIV, by their respective prices as shown in Table XIV-A. The resulting annual totals may be taken provisionally as an index of value production of raw materials, and may be reduced to any desired base. Such indexes are not, however, of much use in themselves, except as they may be employed in connection with the computation of physical production and price indexes.

If the total value production for a given year, as measured by an adequate index, shows an increase over the preceding year, this increase may evidently be attributed either to a growing volume of physical production, or to rising prices, or to a combination of both factors. It therefore follows that suitable indexes of general prices and of physical production, multiplied across year by year, must give an index of value production. This fundamental principle may be expressed by the formula: [1]

$$P_n Q_n = V_n$$

in which

$P_n =$ the price index for a given year

$Q_n =$ the physical production index for the same year

$V_n =$ the value production index for the same year

The formula may also be written:

$$P_n = \frac{V_n}{Q_n}, \text{ and } Q_n = \frac{V_n}{P_n}$$

**An Index of Physical Production.**   The statistical

---

[1] In applying this formula, it is preferable that the two given indexes should be reduced to the same base, but this is not mathematically essential. It is not the absolute value of the derived index numbers that is important, but only their ratios to each other.

TABLE XIV [1]

# PRODUCTION OF SPECIFIED COMMODITIES, U. S., 1870-1920

| Year | Wheat (millions of bushels) | Corn (millions of bushels) | Cotton (millions of bales) | Pig Iron (millions of tons) | Copper (millions of pounds) |
|---|---|---|---|---|---|
| 1870 | 236 | 1,094 | 4.352 | 1.665 | 28 |
| 1871 | 231 | 992 | 2.974 | 1.707 | 29 |
| 1872 | 250 | 1,093 | 3.931 | 2.549 | 28 |
| 1873 | 281 | 932 | 4.170 | 2.561 | 35 |
| 1874 | 308 | 850 | 3.833 | 2.401 | 39 |
| 1875 | 292 | 1,321 | 4.632 | 2.024 | 40 |
| 1876 | 289 | 1,284 | 4.474 | 1.869 | 43 |
| 1877 | 364 | 1,343 | 4.774 | 2.067 | 47 |
| 1878 | 420 | 1,388 | 5.074 | 2.301 | 48 |
| 1879 | 449 | 1,548 | 5.755 | 2.742 | 52 |
| 1880 | 499 | 1,717 | 6.606 | 3.835 | 60 |
| 1881 | 383 | 1,195 | 5.456 | 4.144 | 72 |
| 1882 | 504 | 1,617 | 6.950 | 4.623 | 91 |
| 1883 | 421 | 1,551 | 5.713 | 4.596 | 116 |
| 1884 | 513 | 1,796 | 5.682 | 4.098 | 145 |
| 1885 | 357 | 1,936 | 6.576 | 4.045 | 166 |
| 1886 | 457 | 1,665 | 6.505 | 5.683 | 158 |
| 1887 | 456 | 1,456 | 7.047 | 6.417 | 181 |
| 1888 | 416 | 1,988 | 6.938 | 6.490 | 226 |
| 1889 | 491 | 2,113 | 7.473 | 7.604 | 227 |
| 1890 | 402 | 1,490 | 8.653 | 9.203 | 260 |
| 1891 | 612 | 2,060 | 9.035 | 8.280 | 284 |
| 1892 | 516 | 1,628 | 6.700 | 9.157 | 345 |
| 1893 | 396 | 1,619 | 7.493 | 7.125 | 329 |
| 1894 | 460 | 1,213 | 9.901 | 6.658 | 354 |
| 1895 | 467 | 2,151 | 7.161 | 9.446 | 381 |
| 1896 | 428 | 2,284 | 8.533 | 8.623 | 460 |
| 1897 | 530 | 1,903 | 10.898 | 9.653 | 494 |
| 1898 | 675 | 1,924 | 11.189 | 11.774 | 527 |
| 1899 | 547 | 2,078 | 9.393 | 13.621 | 569 |
| 1900 | 522 | 2,105 | 10.102 | 13.789 | 606 |
| 1901 | 748 | 1,523 | 8.583 | 16.878 | 602 |
| 1902 | 670 | 2,524 | 10.588 | 17.821 | 660 |
| 1903 | 638 | 2,244 | 9.820 | 18.009 | 698 |
| 1904 | 552 | 2,467 | 13.451 | 16.497 | 813 |
| 1905 | 693 | 2,708 | 10.495 | 22.992 | 889 |
| 1906 | 735 | 2,927 | 12.983 | 25.307 | 918 |
| 1907 | 634 | 2,592 | 11.058 | 25.781 | 869 |
| 1908 | 665 | 2,669 | 13.086 | 15.936 | 943 |
| 1909 | 737 | 2,772 | 10.073 | 25.795 | 1,093 |
| 1910 | 635 | 2,886 | 11.568 | 27.304 | 1,080 |
| 1911 | 621 | 2,531 | 15.553 | 23.650 | 1,097 |
| 1912 | 730 | 3,125 | 13.489 | 29.727 | 1,243 |
| 1913 | 763 | 2,447 | 13.983 | 30.966 | 1,224 |
| 1914 | 891 | 2,673 | 15.906 | 23.332 | 1,150 |
| 1915 | 1,026 | 2,995 | 11.068 | 29.916 | 1,388 |
| 1916 | 636 | 2,567 | 11.364 | 39.435 | 1,928 |
| 1917 | 637 | 3,065 | 11.302 | 38.621 | 1,890 |
| 1918 | 921 | 2,503 | 12.041 | 39.055 | 1,994 |
| 1919 | 941 | 2,917 | 11.421 | 31.015 | 1,289 |
| 1920 | 787 | 3,232 | 13.366 | 36.415 | 1,345 |

[1] This table and the one following are taken with some modifications from Babson's *Business Barometers*, by permission of the author.

# TABLE XIV—A

## AVERAGE PRICES OF SPECIFIED COMMODITIES, U. S., (EASTERN MARKETS), 1870-1920

| YEAR | WHEAT PER BU. | CORN PER BU. | COTTON PER BALE | PIG IRON PER TON | COPPER PER LB. |
|------|------|------|------|------|------|
| 1870 | 1.30 | 1.02 | 119.50 | 33.23 | 0.211 |
| 1871 | 1.60 | .77 | 84.50 | 35.08 | .241 |
| 1872 | 1.62 | .70 | 110.50 | 48.94 | .355 |
| 1873 | 1.76 | .63 | 100.50 | 42.79 | .280 |
| 1874 | 1.39 | .86 | 89.50 | 30.19 | .220 |
| 1875 | 1.33 | .84 | 77.00 | 25.53 | .226 |
| 1876 | 1.35 | .628 | 64.50 | 20.75 | .210 |
| 1877 | 1.63 | .593 | 59. | 19.25 | .190 |
| 1878 | 1.24 | .535 | 56. | 17.05 | .165 |
| 1879 | 1.24 | .47 | 54. | 22.82 | .186 |
| 1880 | 1.30 | .55 | 57.50 | 29.86 | .214 |
| 1881 | 1.30 | .62 | 60. | 22.54 | .181 |
| 1882 | 1.32 | .77 | 57.50 | 23.20 | .191 |
| 1883 | 1.17 | .64 | 59. | 19.62 | .165 |
| 1884 | 1.00 | .615 | 54.50 | 16.80 | .110 |
| 1885 | .94 | .51 | 52. | 15.20 | .108 |
| 1886 | .888 | .52 | 46. | 16.77 | .110 |
| 1887 | .88 | .488 | 51. | 20.05 | .138 |
| 1888 | .94 | .593 | 50. | 16.82 | .167 |
| 1889 | .91 | .438 | 53. | 14.35 | .134 |
| 1890 | .92 | .485 | 55. | 15.10 | .156 |
| 1891 | 1.05 | .675 | 43. | 13.78 | .127 |
| 1892 | .908 | .54 | 38.50 | 12.74 | .115 |
| 1893 | .739 | .499 | 42.50 | 11.42 | .107 |
| 1894 | .611 | .509 | 34.50 | 9.93 | .095 |
| 1895 | .669 | .477 | 37. | 10.86 | .105 |
| 1896 | .781 | .340 | 39.50 | 10.29 | .109 |
| 1897 | .954 | .319 | 35. | 9.42 | .113 |
| 1898 | .952 | .376 | 29.50 | 9.46 | .120 |
| 1899 | .794 | .413 | 34. | 16.58 | .177 |
| 1900 | .804 | .453 | 46. | 17.04 | .166 |
| 1901 | .803 | .567 | 43.50 | 13.61 | .161 |
| 1902 | .836 | .684 | 45. | 20.00 | .116 |
| 1903 | .853 | .572 | 55.50 | 17.08 | .132 |
| 1904 | 1.107 | .594 | 58.50 | 12.73 | .128 |
| 1905 | 1.028 | .593 | 49. | 15.57 | .156 |
| 1906 | .865 | .560 | 57.50 | 16.70 | .193 |
| 1907 | .963 | .640 | 60.50 | 23.10 | .200 |
| 1908 | 1.049 | .786 | 53. | 15.54 | .132 |
| 1909 | 1.263 | .767 | 63. | 16.12 | .131 |
| 1910 | 1.118 | .668 | 75.50 | 15.16 | .129 |
| 1911 | .963 | .711 | 65. | 13.67 | .125 |
| 1912 | 1.091 | .711 | 57.50 | 14.93 | .164 |
| 1913 | 1.041 | .711 | 64. | 14.90 | .155 |
| 1914 | 1.094 | .793 | 55.50 | 13.41 | .133 |
| 1915 | 1.291 | .837 | 50.50 | 13.58 | .174 |
| 1916 | 1.468 | .929 | 72. | 18.67 | .272 |
| 1917 | 2.346 | 1.776 | 117.50 | 40.07 | .272 |
| 1918 | 2.31 | 1.840 | 158.50 | 36.52 | .247 |
| 1919 | 2.34 | 1.771 | 161.50 | 32.16 | .192 |
| 1920 | 2.65 | 1.669 | 173. | 44.03 | .175 |

process of developing an index of physical production
may be illustrated by the use of the limited data shown
in Table XIV. In aggregating bushels, bales, tons,
and pounds for any given year, it will be hardly admis-
sible to add the units as tabulated. While the numbers
might be taken abstractly and thus combined into an
index, yet such a process would give as much weight
to a bushel or a pound as to a bale or a ton. It is true
that we might here reduce all our units to pounds, but
even if this were done a pound of copper should be
stressed more than a pound of corn or of iron, because
of its greater importance in the markets. The simplest
way to give each item of production its proper place in
the total will be to remeasure it in terms of a standard
value. That is, we may take as the physical unit the
amount of each commodity that can be bought for a
dollar at a standard price. The number of such units
for each year may then be summated as an index of
physical production.[1]  Expressed algebraically, the
process is:

$$Q_n = \Sigma p_m q_n$$

In so far as the complete index is concerned, this is
equivalent to averaging the quantities as originally
tabulated, weighting them for standard price ($p_m$).

**A Standard Price.** The term ''standard price'' has
been used to imply a price which may be regarded as
representative of a given commodity for the whole in-

[1] It can be shown, as follows, that the product of price and quantity
can properly be taken as physical units:
Let $p$ = price per pound in dollars
and $n$ = number of pounds in a given output.
Then $1/p$ = number of pounds purchasable for one dollar (the new
physical unit).
and $n \div 1/p$ = $np$, the number of new physical units in the output.

terval under consideration. Since the standard price is to be used virtually as a weight, it need not be determined with very great accuracy. But from the point of view of the "theory of errors," it should be the average price during the whole interval of time which is covered by the series of quantity indexes dependent upon it. It follows, therefore, that in periods of rapidly changing prices, somewhat different comparative results may be obtained by varying the interval of time over which the comparisons are made. This may seem anomalous, but it arises inevitably from the fact that the concept of aggregate quantity requires a unit dependent upon value; namely, a dollar's worth. Since value is unstable, the unit of quantity is also unstable. In practice, however, this instability is usually insignificant.

The application of the foregoing principles to the data of Table XIV requires the finding of a standard price for each of the commodities tabulated. As a base for this price, a period of about twenty years prior to the war has been chosen, the purpose being to exclude war prices as extreme, and to emphasize the later rather than the earlier part of the half century studied.[1]   The averages have been taken approximately, and have been slightly modified by the use of data not here cited. They are as follows:

[1] Because of the shifting of relative prices, it might be theoretically preferable to subdivide the half century under consideration into at least three different periods (e. g., 1870-1896, 1896-1914, and 1914-1920), and to derive different sets of weights for use in each period. The series of indexes could be brought together in the overlapping years by a simple adjustment of the bases. But the slight gain in accuracy thus made would not be worth while here, since in any case the results obtained from such meager data cannot be considered as anything more than approximations.

Wheat, per bushel................ $1.04
Corn, per bushel................. .64
Cotton, per bale................. 56.00
Pig Iron, per ton................ 16.00
Copper, per pound................ .16

In applying these prices, it should be remembered that they are in effect weights, and so may be treated as a multiple ratio. They may therefore be changed by division into the more convenient form 13:8:700:200:2.

**Physical Production in the United States.** After the standard prices have been decided upon, each annual item of production is multiplied by its appropriate weight. Two sub-totals are taken for each year, one for crops and one for minerals. These sub-totals constitute two provisional index series. The completed indexes are reduced to a base of 1913. A further step may be taken by noting the fact, as shown in various statistical studies, that an index of mineral production runs very close to an index of manufactures. By suitable weighting, the indexes of crops and minerals may therefore be combined so as to include, in effect, an estimate for the value added by manufacturing. By a comparison of aggregate values and a little experimentation, the requisite weights may be placed at six for crops and four for minerals, manufactures being theoretically included in the latter. This process of weighting and combining is admittedly very crude, and would not be expected ordinarily to give more than a rough approximation to a comprehensive index. But it happens in this case that the five commodities on which the work is based are very dependable and typical as far as production is concerned. The index

# TABLE XV

## INDEXES OF PHYSICAL PRODUCTION, UNITED STATES, 1870-1920

| YEAR | CROPS | MINERALS | GENERAL | |
|------|-------|----------|-----------|------------|
| | | | AGGREGATE | PER CAPITA |
| 1870 | 38 | 5 | 25 | 61 |
| 1871 | 33 | 5 | 22 | 53 |
| 1872 | 38 | 7 | 25 | 60 |
| 1873 | 36 | 7 | 24 | 56 |
| 1874 | 34 | 6 | 23 | 52 |
| 1875 | 45 | 6 | 29 | 64 |
| 1876 | 44 | 5 | 28 | 60 |
| 1877 | 48 | 6 | 31 | 65 |
| 1878 | 51 | 6 | 33 | 67 |
| 1879 | 57 | 8 | 37 | 73 |
| 1880 | 63 | 10 | 42 | 81 |
| 1881 | 47 | 11 | 33 | 61 |
| 1882 | 62 | 13 | 42 | 78 |
| 1883 | 56 | 13 | 39 | 69 |
| 1884 | 64 | 13 | 43 | 76 |
| 1885 | 63 | 13 | 43 | 74 |
| 1886 | 61 | 17 | 43 | 72 |
| 1887 | 57 | 19 | 42 | 69 |
| 1888 | 67 | 20 | 48 | 77 |
| 1889 | 73 | 23 | 53 | 83 |
| 1890 | 59 | 27 | 46 | 71 |
| 1891 | 78 | 26 | 57 | 86 |
| 1892 | 62 | 29 | 49 | 72 |
| 1893 | 59 | 24 | 45 | 66 |
| 1894 | 58 | 24 | 44 | 63 |
| 1895 | 72 | 31 | 55 | 77 |
| 1896 | 76 | 31 | 58 | 79 |
| 1897 | 76 | 34 | 59 | 79 |
| 1898 | 81 | 39 | 65 | 85 |
| 1899 | 77 | 45 | 64 | 83 |
| 1900 | 78 | 46 | 65 | 83 |
| 1901 | 71 | 51 | 63 | 78 |
| 1902 | 92 | 57 | 78 | 95 |
| 1903 | 84 | 58 | 74 | 88 |
| 1904 | 92 | 57 | 78 | 91 |
| 1905 | 97 | 74 | 88 | 100 |
| 1906 | 107 | 80 | 96 | 108 |
| 1907 | 93 | 80 | 88 | 97 |
| 1908 | 100 | 59 | 83 | 90 |
| 1909 | 99 | 85 | 93 | 99 |
| 1910 | 100 | 88 | 96 | 100 |
| 1911 | 100 | 80 | 92 | 95 |
| 1912 | 112 | 98 | 106 | 108 |
| 1913 | 100 | 100 | 100 | 100 |
| 1914 | 112 | 81 | 100 | 98 |
| 1915 | 115 | 101 | 109 | 106 |
| 1916 | 94 | 136 | 110 | 106 |
| 1917 | 104 | 133 | 115 | 109 |
| 1918 | 102 | 137 | 116 | 108 |
| 1919 | 111 | 102 | 107 | 99 |
| 1920 | 115 | 110 | 113 | 103 |

obtained from them in fact checks remarkably well with Stewart's (1890-1920) and Day's (1899-1920) indexes of physical production. The results, put into index form, are shown in Table XV.

**Price Indexes.** The theory of price indexes [1] may be advantageously reviewed in the light of the principles discussed in this chapter. It will be apparent that a price index, to be theoretically precise, must conform to the equation,

$$P_n = \frac{V_n}{Q_n}$$

Since this equation is valid whether the indexes substituted in it have been reduced to a common base or expressed merely in dollars and "dollar's worths," it may be written, by the use of formulas previously considered:

$$P_n = \frac{\Sigma p_n q_n}{\Sigma p_m q_n}$$

The index numbers obtained by the direct use of this formula will have as their base approximately an average index, as determined by the use of average prices in the denominator. If, however, the indexes of value and quantity have first been reduced to a given base, then the price indexes will have that base.

[1] The theories of price and quantity indexes discussed in this book involve the use of the same list of items from one date to another. If a change in the number or character of the items is to be made, as the change from 22 to 43 items in the Bureau of Labor Statistics' index of retail food prices, it is assumed that a new series is begun, and is connected with the old merely by an adjustment of the new base so as to make the indexes at the two overlapping dates agree. The difficult theoretical problem of constructing an index based upon continually shifting lists of commodities is not considered, since it has little immediate practical value. Production indexes, however, to be quite accurate, ought to be gradually broadened to take account of the growing diversification of industry. Hence data such as freight and canal tonnage, which reflect this increasing diversification, are useful in measuring production.

For the purpose of making a comparison with the proportional expenditure method, the formula may be modified by twice inserting the factor $p_m$ in such a way that it will cancel from each term of the numerator summation, as follows:

$$P_n = \frac{\Sigma \dfrac{p_n}{p_m} \times p_m q_n}{\Sigma p_m q_n}$$

The formula as thus written indicates relatives based upon standard prices, and an averaging of these relatives with weights consisting of contemporary physical quantities measured in units of a dollar's worth at standard prices. Briefly stated, the formula means relatives weighted for ''dollar's worths.'' Of course in actual work the simpler form previously stated would be used; the latter form is given merely for comparison and interpretation. Since this method of finding a price index involves a standardization of quantity units throughout a given period, it may be called the method of standard quantities.

**An Approximate Method.** The averaging of prices by the use of quantity weights suggests an approximate method for finding price indexes which may here be stated by way of further comparison. This method uses actual rather than relative prices, and is weighted by the use of the average physical quantities ($q_m$) produced, consumed, or traded during the period of time in question. It is analogous to the method used in finding quantity indexes. Its formula is:

$$P_n = \Sigma p_n q_m$$

The series of indexes so derived may be reduced to any desired base. The formula is a convenient one, but

it is not theoretically valid, and therefore will not give very dependable results. It fails to meet the test implied in the formula:

$$P_nQ_n = V_n$$

**Theoretical Difficulties.** Certain theoretical difficulties encountered in developing a method of finding a price index may be revealed by approaching the problem from another angle. Suppose, first, that we had to deal merely with one commodity having a value for a given year of four dollars, and for the succeeding year of five dollars. If the first year is taken as a base, the price index for the second year is 125. If the second year is taken as a base, the price index for the first year is 80. These two results are consistent, as may be shown by the fact that 80% times 125% is unity, or by the fact that the reciprocal of the index 80 is the index 125.

Let us now consider an analogous problem involving two commodities, with quantities and prices stated, as follows:

<div align="center">

FIRST YEAR

|  | (q) | (p) | (v) |
|---|---|---|---|
| Commodity A, | 10 lbs. @ | $4 = | $40 |
| Commodity B, | 3 bu. @ | 7 = | 21 |
| Value index |  |  | 61 |

SECOND YEAR

|  | (q) | (p) | (v) |
|---|---|---|---|
| Commodity A, | 12 lbs. @ | $5 = | $60 |
| Commodity B, | 2 bu. @ | 6 = | 12 |
| Value index |  |  | 72 |

</div>

Instead of making use of average prices as standards in computing quantities, we shall follow a common usage and take the prices of the base year. Considering the first year as the base, and its prices as the standards for finding quantity units (dollar's worths), we have 61 as both the value and the quantity index. The price index, $V \div Q$, is therefore 100, as it should be in the base year. In the second year $V = 72$ and $Q = 62$, the latter result being obtained by summating the quantities of the second year at the standard prices of the preceding year. The price index for the second year is therefore $72\% \div 62\%$, or $116\%$, as compared with $100\%$ for the first year.

Reversing the computation, we now take the second year as the base year, and its prices as the standard prices. The price index for the second year will now come to 100, since the value index is identical with the quantity index. For the first year the value index will be 61 as before; and the quantity index, obtained by summating the quantities at the standard prices of the second year, will be 68. The price index for the first year is therefore $61\% \div 68\%$, or $90\%$.

Are our two results, obtained from different bases, consistent with each other, as they were when only one price index was used? If so, the product of the two indexes, $116\%$ and $90\%$, should be unity. But their product is actually $104\%$. The other test is that the reciprocal of the index $90\%$ should equal the index $116\%$. But actually the reciprocal of $90\%$ is $111\%$.

It is obvious that the lack of consistency arises from the failure to standardize quantity units by the use of a constant price. If, now, we apply the method of

standard quantities, we obtain the following results for
the two years, respectively:

$$P_1 = \frac{61}{65} = 94 \qquad\qquad P_2 = \frac{72}{67} = 107$$

Considering the first year as a base, we obtain an in-
dex of 114 for the second year.  If the second year is
taken as the base, a consistent result will necessarily
be obtained.  It will be seen that the index for the
second year thus obtained is nearly midway between
the two indexes, 116 and 111, obtained before.

**Fisher's Index.**  Thus the standard quantity method
may be used to avoid the inconsistency encountered
when different standard prices are used in measuring
quantities.  But Professor Fisher has proposed an-
other solution which he regards as the theoretical ideal.
His method involves the finding of the geometrical
average of the two inconsistent results, as seen in the
two indexes, 116 and 111.  This method will here give
a result of 114, which is equal to that obtained by
the method of standard quantities, though as a rule
the results are only approximately equal.  The latter
method has a distinct advantage in that it will give a
consistent series of index numbers for a required inter-
val of time, and the series may be reduced to any de-
sired base. The weight of authority, however, supports
Professor Fisher's solution as being theoretically the
best.  His formula, which is somewhat too complex for
ordinary use, is as follows:

$$P_n = \sqrt{\frac{\Sigma p_n q_n}{\Sigma p_o q_n} \times \frac{\Sigma p_n q_o}{\Sigma p_o q_o}}$$

in which the subscripts $o$ and $n$ indicate a base year
and a relative year, respectively.  The development of

the formula is sufficiently explained in the foregoing discussion. The corresponding formula for quantities ($Q_n$) may be written by interchanging each $p$ and $q$ in the price formula.

It will hardly be profitable to attempt to follow any further the theoretical problems connected with the compiling of price and quantity indexes. It must be emphasized that at present the more significant practical problems relate to the securing of adequate and reliable data rather than to the precision of the methods used in their elaboration. The choice of method, as a rule, will not involve the danger of serious error; while work may be quite invalidated by deficiencies and inaccuracies in the data. An understanding of the theory of the subject will nevertheless be found to have a practical value in the planning of work and in the evaluation of results.

**Quantity Theory Indexes.** An interesting application of price and production indexes to the measurement of general business changes may be made on the basis of the quantity theory of money. This theory in its simplest form states that prices (P) change directly as the circulating medium (M) and its rate of circulation (R), and inversely as the number of physical units traded (N). This statement is expressed algebraically as the following equation of exchange:

$$P = \frac{M\ R}{N}.$$

The equation may be elaborated by subdividing the circulating medium into its various elements, as money and credit, and dealing with each on the basis of its own average rate of circulation. But as it is difficult to

obtain accurate data bearing upon circulation rates, such elaboration will not be attempted here. It should also be noted concerning the quantity theory that there has been much argument as to its validity. But the arguments have not been concerned primarily with the statistical aspects of the subject, to which little objection has been made. They have dealt, rather, with the origin of changes in the equilibrium expressed by the equation.

In order to make a certain provisional use of the formula, further data must be adduced. Values for the term $M$ may be obtained from a study made by Professor E. W. Kemmerer, entitled, *High Prices and Deflation,* which gives estimates of the money and bank credit in circulation during the war period. Brought down to 1920, these estimates are as follows:

|  | Circulation (Millions of dollars) | |
| --- | --- | --- |
| Year | Money | Deposits |
| 1913.............. | 3,390 | 12,678 |
| 1914.............. | 3,505 | 13,430 |
| 1915.............. | 3,682 | 14,411 |
| 1916.............. | 4,159 | 17,840 |
| 1917.............. | 4,914 | 21,273 |
| 1918.............. | 5,579 | 23,771 |
| 1919.............. | 5,793 | 27,928 |
| 1920.............. | 6,060 | 30,300 |

In combining money and deposits into a single index of circulating medium, it is necessary to multiply deposits by two because they circulate, in the form of checks, about twice as fast as money. The numbers

obtained by addition may then be reduced to a 1913 base. For the value of $N$, index numbers of physical production will here be substituted. It is assumed that this may be done because, as a rule, the volume of trade parallels the volume of production. For the value of $P$, the Bureau of Labor Statistics' index of wholesale prices will be used.

An index of the rate of currency circulation may now be readily obtained algebraically by the use of the equation of exchange expressed in the form

$$R = \frac{P \ N}{M}.$$

The resulting index will not, of course, register at all accurately the percentage changes in the rate of currency circulation, owing to the fact that changes in the rate of circulation of goods have been omitted by the substitution of volume of production for volume of trade. But it should show changes in the activity of business by registering the increases or decreases of monetary circulation relative to corresponding changes in the circulation of goods. That is, it should indicate the relative activity of the consumption market. Speculative flurries, which increase the circulation of both currency and securities, will therefore not register at all. That the index does, in fact, show changes in the activity of business from year to year with some degree of certainty, is apparent from the figures. The depression of 1914, the war activity of 1917, the temporary slump of 1919, and the boom culminating in 1920 are all indicated.

The indexes entering into the equation of exchange as just discussed are given below. Taken together,

they present an epitome of the general movement of business. Prices vary directly with purchasing power, $MR$ (demand), and inversely with production, $N$ (supply). It should be observed that the values of $R$ and $N$ here used are mere approximations. They are probably too high in 1914 and 1915, and too low in 1916.

| Year | (P) | (M) | (R) | (N) |
|------|-----|-----|-----|-----|
| 1913 | 100 | 100 | 100 | 100 |
| 1914 | 100 | 106 | 94  | 100 |
| 1915 | 101 | 113 | 97  | 109 |
| 1916 | 124 | 139 | 98  | 110 |
| 1917 | 176 | 165 | 123 | 115 |
| 1918 | 196 | 185 | 123 | 116 |
| 1919 | 212 | 214 | 106 | 107 |
| 1920 | 243 | 232 | 118 | 113 |

A complete statistical verification of the equation of exchange would necessitate ampler data and more extended computations. The point of greatest difficulty would be found to be the construction of an index of the physical volume of trade (N). Theoretically this should include all wealth and services traded in a given period of time, measured in physical units of "dollar's worths" at average prices for the whole interval over which the computation extends. Since $MR$ is equal to the value of the same quantities at current prices, the equation of exchange reduces to the formula for finding price index numbers by the standard quantities method; thus:

$$P_n = \frac{MR}{N} = \frac{\Sigma p_n q_n}{\Sigma p_m q_n}$$

Of course in actual work a process of sampling must necessarily be employed, and various methods for ag-

gregating quantities may be resorted to. The price index thus obtained would theoretically vary somewhat from the usual type of index, because the quantities indicated in the formula refer to trade rather than to production or consumption. Hence goods in which there is unusually active speculation will in effect be weighted heavily.

For discussions of the quantity theory and statistical verifications of the equation of exchange the student may profitably consult Kemmerer, *Money and Credit Instruments in their Relation to General Prices;* Fisher, *The Purchasing Power of Money;* and articles published by King in the weekly service of the Bankers' Statistics Corporation (1920).

**Measurement of the National Income.** We have thus far considered production indexes principally with reference to their use in connection with price indexes. But they are also of direct significance in that they indicate changes in the total national income. By national income is meant the sum of all consumption values in economic goods, services, and property usances, together with increases in stores of goods and extensions of property. This conception differs slightly from aggregate income as privately reckoned, inasmuch as the latter includes increases in capitalized values, particularly of land.

While production indexes show changes in the national income, they serve merely as ratios unless supplemented by estimates of the total income for one or more years. Several such studies have been attempted, but the most complete and authoritative study is one recently made by the National Bureau of Economic

Research. This Bureau is a private organization chartered in 1920 for the purpose of conducting statistical investigations into subjects affecting the public welfare. It is controlled by a board of nineteen directors representing various points of view and interests. Its staff consists of Wesley C. Mitchell, Willford I. King, Frederick R. Macauley, and Oswald W. Knauth. This organization has recently issued a monograph containing estimates of the national income for each of the years 1909-1918, inclusive. The monograph is worthy of careful study, not only for its conclusions, but also as an excellent example of applied statistical methods. The estimates of income were made independently on two different bases; first, by sources of production, and second by incomes received. The two sets of results agree very closely, the maximum difference being 6.9% in 1913. They were averaged to give the final estimates, which were reduced to terms of 1913 prices. As thus reduced, the indexes were as follows, the income for 1913 being 34.4 billion dollars:

| Year | Index |
|------|-------|
| 1909 | 88    |
| 1910 | 94    |
| 1911 | 92    |
| 1912 | 97    |
| 1913 | 100   |
| 1914 | 96    |
| 1915 | 102   |
| 1916 | 118   |
| 1917 | 119   |
| 1918 | 113   |

**The Distribution of Income.** The National Bureau of Economic Research has also made a study of the individual distribution of income in 1918, based in

large part on income tax returns.  The total income for that year was found to be about 58 billion dollars, and the number of personal income recipients was 37.6 million, exclusive of men in active service in respect to both items.[1]  The average income was $1,543; the mode, $957; and the three quartiles were $833, $1,140, and $1,574, respectively.  The distribution is summarized in the following derived tables:

FREQUENCY DISTRIBUTION OF NATIONAL INCOME IN PER-CENTAGES OF CIVILIAN INCOME RECIPIENTS, U. S., 1918

| Income class | Per cent |
|---|---|
| Under $400 | 2.84 |
| $ 400- 800 | 19.51 |
| 800-1200 | 32.16 |
| 1200-1600 | 21.53 |
| 1600-2000 | 9.88 |
| 2000-2400 | 4.64 |
| 2400-2800 | 2.63 |
| 2800-3200 | 1.61 |
| 3200-3600 | 1.07 |
| 3600-4000 | .74 |
| 4000 & over | 3.39 |

CUMULATIVE PERCENTAGES OF CIVILIAN INCOME RECIPI-ENTS AND OF NATIONAL INCOME, U. S., 1918

| Recipients of incomes | Aggregate incomes |
|---|---|
| 10 | 2.7 |
| 20 | 7.2 |
| 30 | 12.5 |
| 40 | 18.7 |
| 50 | 26.0 |
| 60 | 33.6 |
| 70 | 42.2 |
| 80 | 52.5 |
| 90 | 65.2 |
| 99 | 86.2 |
| 100 | 100.0 |

[1] When all are included, the average income becomes $1490.

The first of these tables may be graphed as a frequency curve, omitting the last class of "$4,000 and over," which if accurately represented should be extended to include incomes of several millions.[1] The second may be graphed as a Lorenz curve, the first column being represented on the base line, and the vertical scale being drawn equal to the horizontal scale. A diagonal drawn upward from left to right—the line of equal distribution—will serve as a basis of comparison. Or, the second table may be more simply graphed by obtaining by subtraction the non-cumulative percentages of aggregate income, and representing them as successive vertical blocks above a base line on which are measured the successive ten per cent groups of income recipients.

**Pareto's Law.** If the percentages in the frequency distribution are summated in reverse order, and are plotted vertically against the corresponding lower class limits on double-logarithmic paper (double cycle on each scale), the so-called Pareto's law of income distribution will be illustrated. This law states

[1] If the frequency distribution of incomes is graphed, as represented, on semi-logarithmic paper, with the number of incomes (income recipients) plotted on the vertical arithmetic scale, and the magnitude of the incomes plotted on the horizontal logarithmic scale, a somewhat regular frequency curve is formed. This indicates a type of distribution which is normally symmetrical when the ratio departures from the geometric mean, rather than the differences from the arithmetic mean, are taken as the basis for measuring the dispersion. In such a distribution the median is therefore approximately the geometric mean of the first and third quartiles, instead of the arithmetic mean.
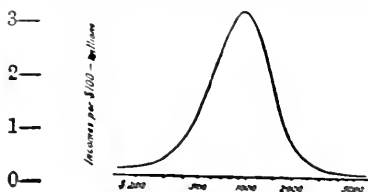
that the curve of income distribution above the mode,
when logarithmically plotted, approximates a straight
line. Much the same result may also be obtained by
transferring the Lorenz curve to double-logarithmic
paper; or it may be more simply evidenced by plotting
the original frequency curve on similar paper. Pa-
reto's law merely calls attention to the type of fre-
quency curve normally followed by income data. This
is a curve which when graphed in the ordinary manner
is seen to be strongly skewed to the right—somewhat
the same type as was suggested by the wage data
studied in the first two chapters. In formulating the
law Pareto thought he had discovered a somewhat in-
flexible and permanent fact of economic relationships.
While he doubtless over-emphasized the inflexibility of
the law, yet he nevertheless pointed out an interesting
illustration of the strong tendency to statistical regu-
larity inherent in biological and social phenomena.

**Income in Other Countries.** The National Bureau
of Economic Research has also made estimates of the
per capita income in several countries for the year
1914. These estimates are based upon prior studies
made by a well-known English authority, Sir Josiah
Stamp. The Bureau's data give the following ratios:

| | |
|---|---|
| United States | 100% |
| Australia | 79 |
| United Kingdom | 73 |
| Canada | 58 |
| France | 55 |
| Germany | 44 |
| Italy | 33 |
| Austria-Hungary | 30 |
| Spain | 16 |
| Japan | 9 |

**The Distribution of Property.** A problem closely related to the distribution of income is the distribution of property ownership. For the United States the best known investigation of the subject is one published by Professor W. I. King in 1915. The total wealth for 1910 was estimated to have been about 200 billion dollars, and its ownership was distributed in a form that is suggested by the curve of income. The richest two per cent of the families were thought to own over fifty per cent of the property. But the data on which such estimates rest involve many uncertainties, and the meaning of the results is further obscured by lack of knowledge of the rate at which fortunes are acquired and dissipated.

## *REFERENCES*

National Bureau of Economic Research, *Income in the United States*, Volume I.

Day, Edmund E., "An Index of the Physical Volume of Production," *Review of Economic Statistics*, September, 1920-January, 1921.

Fisher, Irving, "The Best Form of Index Numbers" (and discussion), *Quarterly Publications of the American Statistical Association*, March, 1921, pp. 533-551.

Fisher, Irving, *The Purchasing Power of Money*.

Hoffman, F. L., "The Economic Progress of the United States During the Last Seventy-five Years," *Quarterly Publications of the American Statistical Association*, December, 1914, pp. 294-318. For further data on the same topic, see Appendix III.

Ingalls, W. R., "Labor the Holder of the Nation's Wealth and Income," *The Annalist*, September 13, 20, and 27, 1920.

King, W. I., *Wealth and Income of the People of the United States*.

Meeker, Royal, "On the Best Form of Index Numbers," *Quarterly Publications of the American Statistical Association*, Sept., 1921, pp. 909-915.

Stewart, Walker W., "An Index Number of Production" (and discussion), *American Economic Review*, March, 1921, pp. 57-81.

Walsh, C. M., *The Measurement of General Exchange Value*.
Working, Holbrook, "What is to be the Future Price Level?"
*The Annalist*, June 27, 1921, p. 686.

## EXERCISES

1. From Tables XIV and XIV-A (pages 76 and 77) find an index of value production (a) for crops and (b) for minerals. Reduce each index to a 1913 base.
2. Divide each item in the index of value production of minerals just obtained, by the corresponding item in the index of physical production of minerals in the United States. Explain the results.
3. Divide each item in the index of value production of crops by the corresponding item in the index of physical production of crops. Explain the results.
4. Plot on semi-logarithmic paper the index obtained in the preceding exercise, together with the index of wholesale prices in the United States. Explain the divergence of the two trends.
5. "The Monthly Review," issued at the Federal Reserve Bank of New York, gives the value of the ten leading crops in the United States at average (standard) prices for recent years as follows:

| Year | | Value (millions of dollars) |
|------|--|------|
| 1910 | ........................ | 5,873 |
| 1911 | ........................ | 5,491 |
| 1912 | ........................ | 6,549 |
| 1913 | ........................ | 5,750 |
| 1914 | ........................ | 6,397 |
| 1915 | . ...................... | 6,831 |
| 1916 | ........................ | 5,907 |
| 1917 | ........................ | 6,507 |
| 1918 | ........................ | 6,418 |
| 1919 | ........................ | 6,626 |
| 1920 | ........................ | 7,284 |
| 1921 | ........................ | 6,118 |

From these data derive an index of agricultural production (base, 1913), and compare it graphically with the corresponding index computed from three leading crops (Table XV, page 81).

6. From the data of crops and prices given below, find index numbers of physical production, value production, and prices, for the year 1914. Express each index in terms of 1913 as a base. (It is assumed that the prices of 1913 represent average prices for a certain period.)

| | Wheat | | Corn | | Cotton | |
|------|------|--------|------|-------|-------|---------|
| Year | Bu. | Price | Bu. | Price | Bales | Price |
| 1913 | 760 | $1.05 | 2450 | $0.72 | 14 | $61.00 |
| 1914 | 890 | 1.10 | 2670 | .80 | 16 | 53.00 |

7. From the following data find indexes of the four terms used in the equation of exchange (quantity theory) for the year 1910, taking the year 1900 as the base.

| | Year: 1900 | 1910 | |
|------|------|------|------|
| Price index ................. | 80 | 100 | |
| Money in circulation........... | 2.2 | 3 | (billions) |
| Deposits subject to check.... | 9.4 | 14.25 | " |
| Physical production ........ | 20 | 30 | " |

8. Plot together on semi-logarithmic paper the index of wholesale prices in the United States for 1890 to 1920 inclusive, and a similar index (base, 1913) of per capita circulation derived from the following data:

| 1890 | $22.82 | 1905 | $31.08 |
|------|--------|------|--------|
| | 23.45 | | 32.32 |
| | 24.60 | | 32.22 |
| | 24.06 | | 34.72 |
| | 24.56 | | 34.93 |
| 1895 | 23.24 | 1910 | 34.33 |
| | 21.44 | | 34.20 |
| | 22.92 | | 34.34 |
| | 25.19 | | 34.56 |
| | 25.62 | | 34.35 |
| 1900 | 26.93 | 1915 | 35.44 |
| | 27.98 | | 39.29 |
| | 28.43 | | 45.74 |
| | 29.42 | | 50.81 |
| | 30.77 | | 54.33 |
| | | 1920 | 57.04 |

9. The two indexes of physical production in the United States (1913-1920) given below, are adapted (a) from Day's and (b) from Stewart's comprehensive studies.

Using data for prices and circulating medium cited in the discussion of the quantity theory of money, derive two indexes for $R$, 1913-1920, and compare them with the corresponding index included in the data just mentioned.

| Year | (A) | (B) |
|------|-----|-----|
| 1913 | 100 | 100 |
| 1914 | 98 | 100 |
| 1915 | 105 | 111 |
| 1916 | 111 | 116 |
| 1917 | 114 | 123 |
| 1918 | 113 | 124 |
| 1919 | 106 | 119 |
| 1920 | 111 | 122 |

# CHAPTER V

## TRENDS AND CYCLES

**The Nature of Trends.** The interpretation of a time series of index numbers usually requires the determination of the trend. A trend is a derived series of index numbers following the general course of the given items, but shortening or eliminating the fluctuations. Its significance is best grasped by means of a graphic representation. An inspection of such graphic work will show that trends may vary from straight lines on the one hand, to curves almost conforming to the given items on the other. Whether a trend is drawn as a straight or an irregular line depends in part on the nature of the data and in part on the purpose to be served.

**The Free-hand Method.** A good draftsman commonly determines trends merely by charting his data and, without any preliminary computations, drawing a straight or curved median line through them.[1] Such work may be done entirely free-hand, or by the use of irregular curves and other drafting material, but in either case it is classed as a free-hand method. For many purposes this method will prove satisfactory, and should be practiced by the student until it can be used with facility. For even though more elaborate

---

[1] If a trend based upon geometric means is desired, the data may be plotted on semi-logarithmic paper.

methods are to be applied in actual work, practice in
the free-hand drawing of trends will be helpful.  It
will develop an appreciation of the requirements of a
given problem, without which the best of mathematical
methods are likely to be misapplied.

An example of the free-hand method of drawing a
trend is given in the graph of the index of real wages,
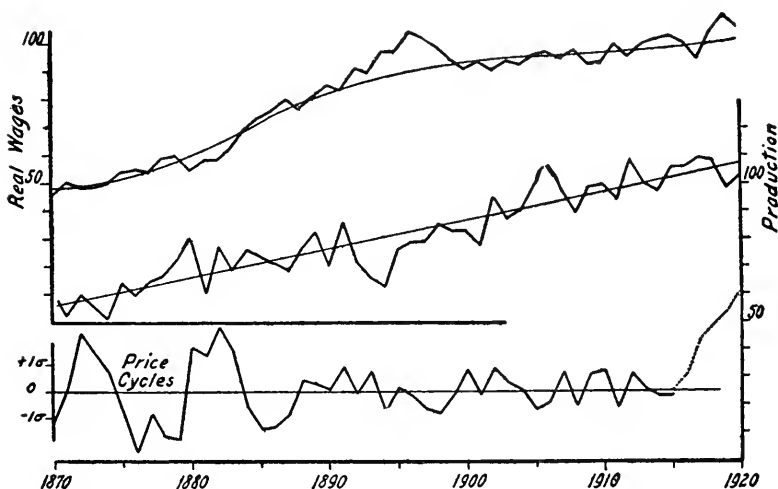


FIGURE 6.   Trends and cycles.    Upper line, index of real wages (hour
rates) in the United States (see Table X) and its trend; middle line,
index of per capita production (see Table XV) and its trend; lower
line, cycles of wholesale prices (see Figure 9).
    Reprinted, with permission, from *The Quarterly Journal of the Univer-
sity of North Dakota*, January, 1922.

in Figure 6.  In this case a more precise method would
not be applicable because of the inaccuracies caused
by the substitution of wholesale prices for the cost of
living in the computation.  As was previously stated,
there is reason to think that the marked rise in real
wages appearing during the decade 1890-1900 is an
exaggeration.  The trend was therefore drawn so as to

discount this rise. As a result it does not conform to the usual rule that the deviations above and below it should balance.

**Method of Semi-Averages.** When a straight-line trend is to be drawn through a fairly long series, the free-hand method may be improved upon by means of a simple calculation. The average of the series may be plotted on the middle ordinate, and the straight-line trend drawn by inspection through this point. The plus and minus deviations must then necessarily balance. Or, better still, the series may be divided into two equal parts and an average taken for each part. These averages may be plotted on the middle ordinate of each half series, respectively, and the trend drawn through the two points. If the series consists of an odd number of items, the middle item and unit of time may be divided between the two parts. This method of semi-averages is the one that was used in drawing the trend of per capita production in Figure 6. The trend thus obtained is, in a long series, nearly identical with the so-called line of least squares to be discussed later.

**The Moving Average.** Among the trends that are found by mathematical methods, perhaps the best known is the moving average. The process of computing the moving average is simple in principle, but it is usually tedious in practice even when calculating machines are used. By this method the position of the trend at any given period in the series is found by averaging a certain number of items centering at that period. Just how many items should be included in the average must be determined by the nature of

the data. If the series shows a cyclic movement of known length, then by taking the number of items covering this length of time, the cycles will be smoothed. If monthly data having a pronounced seasonal swing are being studied through an interval of several years, a twelve-month moving average is appropriate.[1] The method of finding the moving average is illustrated in the following table:

| Year | Per Capita Production | Moving Averages 3-Year | Moving Averages 5-Year |
|------|------|------|------|
| 1910 | 100 | | |
| 1911 | 95 | 101 | |
| 1912 | 108 | 101 | 100 |
| 1913 | 100 | 102 | 101 |
| 1914 | 98 | 101 | 104 |
| 1915 | 106 | 103 | 104 |
| 1916 | 106 | 107 | 105 |
| 1917 | 109 | 108 | 106 |
| 1918 | 108 | 105 | 105 |
| 1919 | 99 | 103 | |
| 1920 | 103 | | |

In explanation of the foregoing table it may be said that the first number in the three-year moving average (101) is obtained by averaging the first, second, and third items (100, 95, and 108). The second number (101) is obtained by averaging the second, third, and fourth items (95, 108, and 100). The results are written to the nearest unit, and are placed opposite the middle one of the three items averaged. In the same way the succeeding averages are derived. The five-year moving average is similarly computed, except

[1] The twelve-month moving average centers between the sixth and seventh month in each computation. In order to make the trend thus obtained fall on the same ordinates as the original items, it is necessary to adjust it by deriving from it a two-month moving average. The deviations of the original items may then be readily obtained.

that five items are averaged at a time. In practice
the work may be somewhat abridged, after the total
of the first group of items is found, by deriving the
next total from it. This may be done by adding to
the first total the difference between the next item
about to be included and the one about to be dropped.
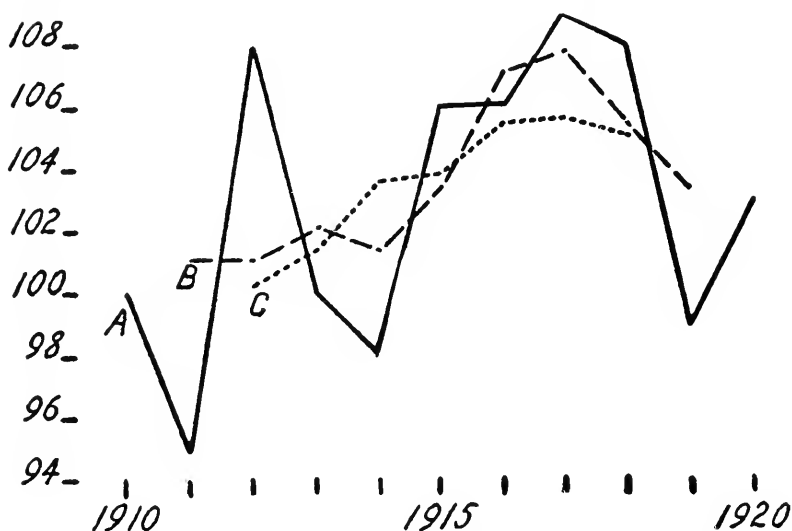


FIGURE 7. Moving averages. A, Index of per capita production,
1910-1920. B, Three-year moving average of same. C, Five-year mov-
ing average of same.

Thus, the first five items above give 501 (average,
100). In obtaining the total of the second to the sixth
items, inclusive, the sixth item (106) will be added
and the first item (100) will be dropped; that is, a
balance of six will be added. The new total is there-
fore 501 + 6, or 507 (average 101). In the same way
the succeeding totals and averages may be derived.[1]

[1] Mathematicians sometimes prefer a more complex form of the
moving average known as the progressive mean. This is similar to

The two moving averages just described, and the index on which they are based, are plotted in Figure 7. The figure will serve to make clear the general rule that the more inclusive the moving average, the smoother the trend will be. A disadvantage of the method will also be observed. The moving average derived from a given series of items will always be shorter than the series; and the more inclusive it is, the shorter it will be. It is possible, however, to find a tentative substitute for the lacking items in the trend by repeating the extreme items in finding the averages. Thus, in the foregoing five-year moving average, a trend item for 1911 might have been obtained as follows:

$$\frac{2 \times 100 + 95 + 108 + 100}{5} = 101$$

and for 1910:

$$\frac{2 \times 100 + 2 \times 95 + 108}{5} = 100$$

This method has the mathematical advantage of making the sum of the trend items equal to the sum of the data—a fact which may be found convenient to use as a basis for checking the computations. It may also be adapted to finding a current trend item for a series of index numbers that is kept up to date.

**The Line of Least Squares.** If a straight-line trend is at all adapted to a given series, the most satisfactory mathematical trend to use is one known as the

the moving average, as illustrated and explained, except that weights are used in taking the average. The weights are derived by the binomial theorem, and are the frequencies of a theoretical curve of distribution. Thus in taking a five-year progressive mean, each group of five terms is averaged by applying to the terms in succession the weights 1:4:6:4:1. Similarly, a seven year progressive mean would make use of the weights 1:6:15:20:15:6:1 (cf. Slichter, *Elementary Mathematical Analysis*, p. 194).

line of least squares. The name is derived from the fact that the line is so drawn that the square of the deviations of the data from it, as measured on the ordinates, is always a minimum. The line of least
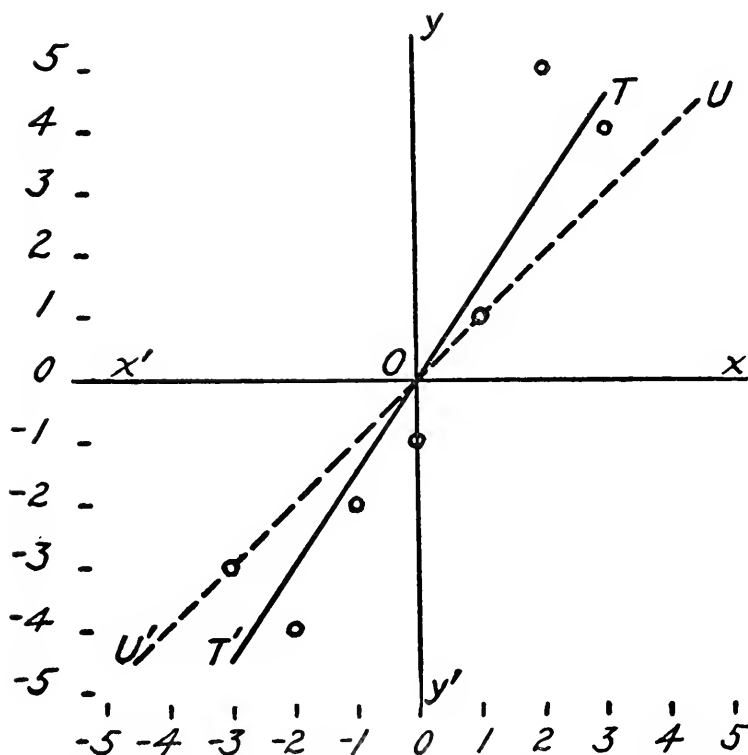


FIGURE 8. Straight-line trend. TT', line of least squares for the seven points indicated; UU', line of unit slope.

squares should be thoroughly understood, not only because of its usefulness as a trend, but also because it is the basis from which the principal method of computing correlation is developed. In explaining it, the following simple illustration will be taken:

Suppose that it is required to find a straight-line trend for the index $y$ (see next page), the average of which is zero. The data are plotted upon coordinate paper—the $x$ and $y$ scales having preferably the same unit—as shown in Figure 8. The average of the data is made to fall on the x-axis, and the middle item is plotted on the y-axis. If we consider the vertical distance of each index from the x-axis to represent a force bearing upon that line, then the total moments of these forces will be expressed by the sum of the $xy$'s. This sum may be compared with the sum of the moments of a line passing through the same ordinates, and forming an angle of forty-five degrees with the $x$ and $y$ axes at their point of intersection.[1] Such a line is said to have a unit slope; that is, it rises one unit $(y)$ for each unit $(x)$ to the right. Its slope may also be expressed by saying that the tangent of its angle (UOX) is unity. The sum of its $xy$'s is, of course, identical with the sum of its $x$'s squared. The slope of the line of least squares is found by comparing the moments of the data with those of the line of unit slope, as expressed in the formula:

$$S = \frac{\Sigma xy}{\Sigma x^2}$$

in which,

$S =$ slope of the line of least squares, or tangent of its angle

$x =$ position of items relative to middle ordinate

$y =$ items, as given

The data $(y)$, their moments $(xy)$, the moments of the line of unit slope $(x^2)$, and the computation of

[1] The angle will vary, of course, if the $x$ and $y$ scales differ.

the line of least squares, are shown in the following
table:

| y | x | $x^2$ | xy | Trend |
|---|---|---|---|---|
| -3 | -3 | 9 | 9 | -4.5 |
| -4 | -2 | 4 | 8 | -3 |
| -2 | -1 | 1 | 2 | -1.5 |
| -1 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 1.5 |
| 5 | 2 | 4 | 10 | 3 |
| 4 | 3 | 9 | 12 | 4.5 |

$$A = 0 \qquad 28 \quad ) \quad 42$$

$$S = 1.5$$

The last column headed "trend" gives the line of least
squares. This column is computed from the middle or-
dinate (0) by adding the slope (S = 1.5) once for each
successive ordinate in a positive direction, and sub-
tracting it once for each successive ordinate in a nega-
tive direction.

Data having an average of zero have here been
taken merely for the sake of simplicity; the same
process with but little modification may be applied to
any values. The dates or other numbers correspond-
ing to the items cannot be used, however, but must
be replaced by an $x$ scale centering at the middle point
of the series. The average of the data is found,
and the trend is computed from this average by suc-
cessive additions of the slope in a positive direction,
and subtractions in a negative direction. The method
is illustrated by the use of the following data, which
parallel those used in the preceding illustration, except
that the average is 100. This increase in the values of

$y$ disappears from $\Sigma xy$ because it affects equally both the minus and the plus items.

| Year | y | x | $x^2$ | xy | Trend | |
|------|------|------|------|------|------|------|
| 1900 | 97 | –3 | 9 | –291 | 95.5 | |
| 1901 | 96 | –2 | 4 | –192 | 97 | |
| 1902 | 98 | –1 | 1 | –98 | 98.5 | |
| 1903 | 99 | 0 | 0 | 0 | 100 | (average) |
| 1904 | 101 | 1 | 1 | 101 | 101.5 | |
| 1905 | 105 | 2 | 4 | 210 | 103 | |
| 1906 | 104 | 3 | 9 | 312 | 104.5 | |
| 7 | )700 | | 28 | ) 42 | | |
| A = 100 | | | | S = 1.5 | | |

The following details may be noted: (a) If there are an even number of ordinates, the y-axis will lie midway between the two middle ordinates, which are numbered as –0.5 and 0.5 respectively. The horizontal positive scale will therefore read 0.5, 1.5, 2.5, etc., and the negative scale will be the reverse. (b) It will sometimes be found that the value of $S$ is negative. This indicates a downward slope of the line of least squares. (c) The position of the line of least squares is described by designating the period coinciding with the $y$ axis as the point of origin, and by expressing the value of $y$ algebraically in terms of the average and the slope. Thus in the above illustration the point of origin is 1903, and the equation of the trend is y = 100 + 1.5x.[1]

[1] It has been suggested that the method of least squares might be applied to the finding of a price index (*Quarterly Journal of Economics*, August, 1921, page 567). The expenditure for any given commodity may be plotted on a coordinate chart as the value of $y$, and the number of units bought may be plotted as the value of $x$. The slant of a line (tangent of the base angle) drawn from the intersection of the axes to the point determined by the values of $x$ and $y$, represents the price. The average price of a number of commodities may be

**The Parabola Trend.** A broad treatment of the subject of curve fitting would lead the student beyond the range of ordinary statistical work. We shall not, therefore, follow the subject farther, except to take up



FIGURE 9. Index of wholesale prices in the United States (see Table X) and trend. Indexes for 1870-1880 converted to a gold basis.

a simple method of adjusting a parabola to an index. The method is one which is often used by engineers, and has also recently come into use to some extent among statisticians.

taken as the slant of the line of least squares determined by all the coordinate pairs of $x$ and $y$, and having the intersection of the axes as the point of origin ($y = Sx$). In such a case, of course, all the values of $xy$ will be positive. To give definite comparisons at different dates, this method would require the use of "dollar's worths" as physical units. While the method is ingenious, it is of questionable validity, since in effect it involves a weighting of the prices by the square of the quantities in the process of finding the average. A defense of the method on the basis of the use of least squares in the theory of errors does not appear to be valid, since the theory of errors would call for merely the arithmetic mean of the number of determinations.

The method of adjusting a parabola will be illustrated by applying it to the Bureau of Labor Statistics' wholesale price index for the years 1896 to 1915 inclusive, as charted in Figure 9. This figure includes also the same index for the years 1870-1895 (gold prices), to which a line of least squares has been fitted. But it may be easily seen that a similar trend would not be suited to the succeeding index numbers. The somewhat regular curve of the latter portion of the index indicates that a parabola of the second order would be appropriate.

In fitting the parabola, the year 1895 has been taken as the point of origin, though this year is not included in the results. It is estimated by inspection of the graph that the trend, if extended to 1895, would have a value of 64 at that date. Two other points determining the trend may be similarly located, one at about the middle of the series and one at the end. These points have been taken as 88 for the year 1905 (x = 10), and 102 for the year 1915 (x = 20). We have, then, these coordinate values of $x$ and $y$:

$$\text{If x} =\ \ 0, \qquad \text{y} =\ \ 64$$
$$\text{If x} = 10, \qquad \text{y} =\ \ 88$$
$$\text{If x} = 20, \qquad \text{y} = 102$$

The equation of a parabola of the second order is,

$$y = a + bx + cx^2$$

If the coordinate values of $x$ and $y$ as just stated are substituted successively in this equation, the following results will be obtained:

$$64 = a$$
$$88 = a + 10b + 100c$$
$$102 = a + 20b + 400c$$

Solving for the constants gives,

$$a = 64$$
$$b = \phantom{0}2.9$$
$$c = \phantom{0}-.05$$

Substituting these values in the original equation gives the equation of the required trend,

$$y = 64 + 2.9x - .05x^2$$

from which the value of each item in the trend may be found by substituting the coordinate value of $x$.

Since trends are used as a basis for measuring fluctuations, the deviations of the data from the trend are usually computed. This is done by subtracting each item in the trend from the corresponding item in the data. If the trend has been accurately constructed, the positive and negative deviations should be practically equal; that is, their sum should be zero. Where the parabola has been used, a certain error will probably be found to have resulted from the fact that the original points determining the curve were located merely by inspection. An adjustment (centering) to remove this error may be made by finding the sum of the deviations ($\Sigma D$), dividing it by the number of the items (N), adding the result $\left(\dfrac{\Sigma D}{N}\right)$ to each of the trend items, and subtracting it from each of the deviations. This is expressed in the equation of the trend simply by adding the correction $\left(\dfrac{\Sigma D}{N}\right)$ to the value of $a$. In work in correlation, however, the correction may be more easily made by another method, as will be evident later. When comparisons of the fluctuations in different series are to be made, either by graphing or

by the computation of a coefficient of correlation, it is often necessary to find the standard deviation. For graphic representation, the deviations are reduced to multiples of the standard deviation, which serves as a comparable unit. Table XVI shows the derivation of

TABLE XVI

DERIVATION OF TREND AND CYCLES OF WHOLESALE PRICES, BASED ON BUREAU OF LABOR STATISTICS INDEX, UNITED STATES, 1896-1915

Equation of trend, $y = 64 + 2.9x - .05x^2$     Point of origin, 1895.
Equation, corrected, $y = 63.825 + 2.9x - .05x^2$

| YEAR | PRICE INDEX | x | TREND y | D | D CENTERED | $D^2$ | CYCLES [1] $D/\sigma$ |
|------|------|------|------|------|------|------|------|
| 1896 | 66 | 1 | 66.85 | −.85 | −.68 | .4624 | −.33 |
| 1897 | 67 | 2 | 69.60 | −2.60 | −2.42 | 5.8564 | −1.17 |
| 1898 | 69 | 3 | 72.25 | −3.25 | −3.08 | 9.4864 | −1.49 |
| 1899 | 74 | 4 | 74.80 | −.80 | −.62 | .3844 | −.30 |
| 1900 | 80 | 5 | 77.25 | 2.75 | 2.92 | 8.5264 | 1.41 |
| 1901 | 79 | 6 | 79.60 | −.60 | −.42 | .1764 | −.20 |
| 1902 | 85 | 7 | 81.85 | 3.15 | 3.32 | 11.0224 | 1.60 |
| 1903 | 85 | 8 | 84.00 | 1.00 | 1.18 | 1.3924 | .57 |
| 1904 | 86 | 9 | 86.05 | −.05 | .12 | .0144 | .06 |
| 1905 | 85 | 10 | 88.00 | −3.00 | −2.82 | 7.9524 | −1.36 |
| 1906 | 88 | 11 | 89.85 | −1.85 | −1.68 | 2.8224 | −.81 |
| 1907 | 94 | 12 | 91.60 | 2.40 | 2.58 | 6.6564 | 1.25 |
| 1908 | 91 | 13 | 93.25 | −2.25 | −2.08 | 4.3264 | −1.00 |
| 1909 | 97 | 14 | 94.80 | 2.20 | 2.38 | 5.6644 | 1.15 |
| 1910 | 99 | 15 | 96.25 | 2.75 | 2.92 | 8.5264 | 1.41 |
| 1911 | 95 | 16 | 97.60 | −2.60 | −2.42 | 5.8564 | −1.17 |
| 1912 | 101 | 17 | 98.85 | 2.15 | 2.32 | 5.3824 | 1.12 |
| 1913 | 100 | 18 | 100.00 | 0 | .18 | .0324 | .09 |
| 1914 | 100 | 19 | 101.05 | −1.05 | −.88 | .7744 | −.43 |
| 1915 | 101 | 20 | 102.00 | −1.00 | −.82 | .6724 | −.40 |

$$\begin{array}{cccc} & 16.40 & 17.92 & 20)85.9880 & 8.66 \\ & -19.90 & -17.92 & & -8.66 \\ \hline & 20)-3.50 & 0. & \sigma^2 = 4.2994 & 0. \\ & K = -.175 & & \sigma = 2.07 \end{array}$$

[1] Any set of deviations taken from an average or trend, and intended to measure cyclic movements, are commonly designated as cycles. They need not necessarily be reduced to units of the standard deviation. In some cases no complete cyclic movement may be discovered, but the same designation may be used.

the trend just discussed, together with the correction, and the reduction of the deviations to multiples of the standard deviation. Figures 10 and 11 show price cycles as thus measured compared with other cycles similarly obtained.

The parabola may be used where compound curves are required by adding the term $dx^3$, and perhaps $ex^4$ to the equation. For each term thus added, an addi-
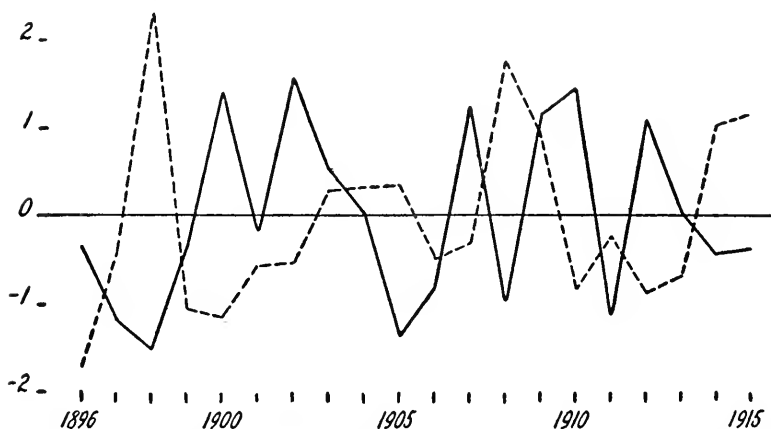


FIGURE 10.   Cycles of wholesale prices (solid line) and commitments to New York State prisons (broken line).
Reprinted, with permission, from *The Quarterly Journal of the University of North Dakota,* January, 1922.

tional point may be located by inspection, and the trend may thus be more exactly fitted. But the work of solving and applying such equations becomes very laborious. With practice, however, the student will find ways of abbreviating the process and modifying it to suit his purposes. Often the terms of the equation may be estimated by simple experimentation. The position of the point of origin may be varied to suit given requirements. A compound curve shaped some-

what like an italic *f* may be obtained by using only odd numbered powers of $x$ in the equation, and taking the point of origin near the middle of the original series. With a little ingenuity, sine curves and other trends may be experimentally fitted.[1]
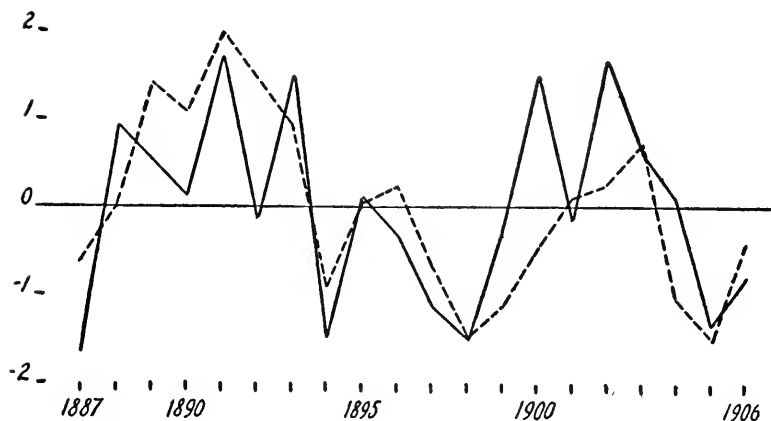


FIGURE 11. Cycles of wholesale prices (solid line) and marriage rate (broken line) in the United States.

Reprinted, with permission, from *The Quarterly Journal of the University of North Dakota*, January, 1922.

**Analyzing Business Barometers.** A somewhat intricate problem in trends is met when monthly data are employed as indexes, or barometers, of business conditions. Since such barometers are very generally consulted as guides to business activities, their inter-

[1] When it appears that the trend of an index series increases or decreases by approximately a fixed ratio, an exponential curve may be readily fitted as follows. Find the logarithms of the data, plot them, and construct a straight-line trend by means of the semi-averages, or by a line of least squares. Read the items of the trend from the chart, consider them to be logarithms, and find the corresponding numbers. These numbers will be the items of the required trend. In a long series the work may be abbreviated by grouping the original items, as by decades, and fitting the curve to the averages of these groups.

pretation is a matter of great practical importance. The difficulty involved in their use lies in the complexity of the influences playing upon them. For convenience of analysis these influences have been classified as (1) a seasonal variation usually due to the dependence of industry upon weather conditions, illustrated by the rising of the interest rate with the movement of crops, (2) a cyclic movement covering an interval of several years and marked by alternating industrial depression and activity, and (3) a secular trend or gradual change due in most cases to the growth movement, as seen in the increase in production. In addition to influences which may be appropriately classed under one of these three headings, there are others that must be looked upon as more or less accidental interruptions, of which no exact account can be taken.

**Measuring Seasonal Variations.** The most satisfactory method of analyzing monthly business barometers is first to compute an index of seasonal variations, and then to subtract it, month by month, from an index of the data based upon the secular trend. The result is an index of the cycles. As has already been noted, the twelve-month moving average is sometimes assumed to measure the data as distinct from the seasonal variations. But such a method of elimination takes into account the fluctuations of only one year at a time. Other methods have therefore been resorted to with the purpose of measuring the seasonal swing more exactly on the basis of several years. A simple method of this sort, which may be considered valid for a period in which the cyclic in-

TABLE XVII

AVERAGE MONTHLY INTEREST RATE, SHORT-TIME LOANS, U. S., 1909-1913, AND COMPUTATION OF SEASONAL VARIATIONS

| MONTH | YEAR | | | | | MONTH TYPE (AVERAGE) | SECULAR TREND | SEASONAL INDEX | D(%) |
|---|---|---|---|---|---|---|---|---|---|
| | 1909 | 1910 | 1911 | 1912 | 1913 | | | | |
| January | 4.5 | 4.8 | 4.3 | 4.0 | 5.0 | 4.52 | 4.702 | 96 | -4 |
| February | 3.7 | 4.5 | 4.0 | 3.7 | 5.0 | 4.18 | 4.721 | 89 | -11 |
| March | 3.7 | 4.7 | 4.0 | 4.1 | 5.6 | 4.42 | 4.740 | 93 | -7 |
| April | 3.8 | 4.8 | 3.6 | 4.4 | 5.7 | 4.46 | 4.759 | 94 | -6 |
| May | 3.8 | 4.8 | 3.6 | 4.1 | 5.4 | 4.34 | 4.778 | 91 | -9 |
| June | 3.9 | 4.9 | 3.6 | 4.0 | 5.9 | 4.46 | 4.797 | 93 | -7 |
| July | 3.6 | 5.2 | 3.9 | 4.5 | 6.0 | 4.64 | 4.816 | 97 | -3 |
| August | 4.4 | 5.7 | 4.4 | 5.2 | 6.0 | 5.14 | 4.835 | 106 | 6 |
| September | 4.6 | 5.8 | 4.5 | 5.8 | 6.0 | 5.34 | 4.854 | 110 | 10 |
| October | 5.4 | 5.8 | 4.5 | 6.0 | 5.8 | 5.50 | 4.873 | 113 | 13 |
| November | 5.5 | 5.7 | 4.0 | 6.0 | 5.7 | 5.38 | 4.892 | 110 | 10 |
| December | 5.7 | 4.5 | 4.5 | 6.0 | 5.8 | 5.30 | 4.911 | 108 | 8 |
| Average | 4.38 | 5.10 | 4.07 | 4.82 | 5.66 | 4.807 | 4.807 | 100 | 0 |

SECULAR TREND—LINE OF LEAST SQUARES

| YEAR | ANNUAL AVERAGE | x | x² | MOMENTS |
|---|---|---|---|---|
| 1909 | 4.38 | -2 | 4 | -8.76 |
| 1910 | 5.10 | -1 | 1 | -5.10 |
| 1911 | 4.07 | 0 | 0 | 0 |
| 1912 | 4.82 | 1 | 1 | 4.82 |
| 1913 | 5.66 | 2 | 4 | 11.32 |
| | | | 10 | 2.28 |

$S = 2.28 \div 10 = .228$ (annual)

$S \div 12 = .019$ (monthly)

fluences are moderate and well distributed along an approximately straight-line trend, is illustrated in Table XVII.

**The Method of Averages.** The problem stated and solved in the table is the finding of the seasonal variations in the interest rate on the somewhat slender basis
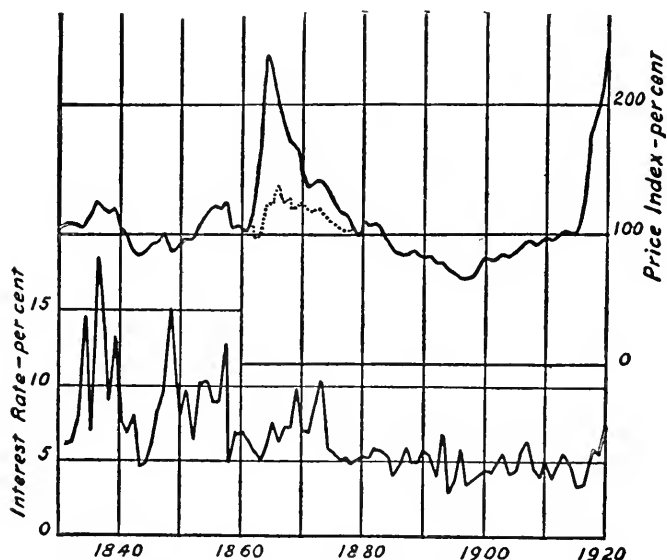


FIGURE 12. Average interest rate on commercial paper in the United States (lower line) compared with index of wholesale prices (upper line).

Adapted from *Monthly Review*, New York Federal Reserve Bank.

of the five years 1909 to 1913, inclusive. The monthly rates as given are first averaged both by columns and rows in order to obtain the annual averages and the month averages. The process may therefore be called the method of averages.

If the interest rate had maintained approximately the same level from year to year during the interval

studied, nothing more would be needed than to reduce the month averages, or types, to a percentage of their own annual average. The results would form an index of the seasonal variations. But if there is a secular movement, it must be canceled from such an index. If an interval of half or three-quarters of a century is taken into account, a general downward trend of the interest rate may be discovered, as may be seen in Figure 12. But the five years here studied happen to be an exception. By means of the line of least squares, a positive slope of 0.019 monthly is revealed. By applying this slope to the average of the month types, an annual trend is constructed, having its point of origin midway between the June and July items. It will be noted that one-half the slope must be added to the average to obtain the July trend item, and the same amount subtracted to obtain the June item. In obtaining the remaining trend items, the slope is applied as previously explained.

Confusion may perhaps here arise from the fact that the slope was computed from the five annual averages, but was applied to the construction of a trend with which to compare the monthly data. But it should be obvious that in this case a slope obtained from the month averages would be materially affected by the seasonal swing. This we wish to retain, while the secular trend we wish to cancel. Of course we could find the secular trend from the monthly data taken consecutively as sixty items, but such a method would be unnecessarily laborious. Hence we find it from the five annual averages, and apply it to the construction of a line that will serve to cancel the secular trend from

the month types. This cancellation is accomplished by dividing the month averages, item by item, by the trend. The quotients should be centered, if necessary, by reducing them to percentages of their common average. The result is an index of seasonal variations, from which the percentage deviations, month by month, may be directly stated.[1]

**Applying the Seasonal Index.** The method of applying the seasonal index has already been suggested, and may be described as follows. A secular trend for the whole period under consideration is constructed—in this case the line of least squares already found may be extended—and the data are reduced month by month to percentages of the trend. From each month's item as thus found the seasonal index for the same month is subtracted. The remainders are assumed to measure the cycles, and may be plotted as deviations above and below a horizontal axis. The seasonal index may, with caution, be applied to other years than those from which it is derived; of course, the greater the number of normal years from which it is derived, the safer such an extension of its use to comparable years becomes.

**The Link-relative Method.** A complex but more accurate method for measuring seasonal variations has

---

[1] Another variation of the method of averages—one that is perhaps theoretically preferable, but which involves more extended calculations—may be briefly described as follows: A twelve-month moving average of the monthly data through a given series of years is first found. This is adjusted to make it conform to the ordinates of the original series by deriving from it a two-month moving average. There is then obtained, for each month, the ratio of the original monthly item to the corresponding adjusted moving average. The median of the ratios so obtained for the Januaries is taken as the index of seasonal variation for January; and index numbers for the other months are similarly obtained. The twelve results are then centered, if neces-

been developed by Professor Persons, and applied to the analyses appearing in the early numbers of the *Review of Economic Statistics*. In a somewhat simplified form, this method is illustrated in Table XVII-A, which is based on the data of Table XVII. Briefly stated, the method consists in finding what are called "link-relatives"; that is, the percentage which the index of each month is of the preceding month. The median link-relatives are then selected from each month's series, and are tabulated as the month types.[1] Beginning with December as a base (100%), the types are multiplied consecutively, producing an index series from January to December for a typical year. If the final December item fails to come out to 100%, in conformity with the base in the preceding December, a secular trend is evidently disturbing the index. The discrepancy, if moderate, may be removed by distributing it throughout the year; that is, by subtracting one-twelfth of it from the January index, two-twelfths from the February index, and so on through the year.[2] The results are designated an adjusted index. The adjusted items are next centered by reducing them to a base of the average of the series. In making the computations the figures were carried to one more place than is shown in the table.

The superiority of this method lies in the fact that in taking the link-relatives, and in selecting their medians as the month types, the effects of the cyclic

sary, by reducing them to percentages of their common average (Cf. Jordan, *Business Forecasting*, p. 212).

[1] To get the best results, the median should be based on a larger number of years than are here taken.

[2] A more exact method is to apportion the discrepancy geometrically; that is, to divide the January index by the twelfth root of the final December index (written as a decimal), the February index by the square of this root, and so on.

TABLE XVII-A

LINK-RELATIVES OF MONTHLY INTEREST RATES (See Table XVII)

| MONTH | YEAR | | | | | MONTH TYPE (MEDIAN) | INDEX | INDEX ADJUSTED | INDEX CENTERED | D (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1909 | 1910 | 1911 | 1912 | 1913 | | | | | |
| January.......... | 94 | 84 | 96 | 89 | 83 | 89 | 89 | 88 | 96 | -4 |
| February......... | 82 | 94 | 93 | 93 | 100 | 93 | 83 | 81 | 89 | -11 |
| March........... | 100 | 104 | 100 | 111 | 112 | 104 | 86 | 84 | 91 | -9 |
| April............ | 103 | 102 | 90 | 107 | 102 | 102 | 88 | 85 | 92 | -8 |
| May............. | 100 | 100 | 100 | 93 | 95 | 100 | 88 | 84 | 91 | -9 |
| June............. | 103 | 102 | 100 | 98 | 109 | 102 | 90 | 85 | 93 | -7 |
| July............. | 92 | 106 | 108 | 112 | 102 | 106 | 95 | 90 | 98 | -2 |
| August........... | 122 | 110 | 113 | 115 | 100 | 113 | 107 | 101 | 110 | 10 |
| September........ | 105 | 102 | 102 | 112 | 100 | 102 | 109 | 103 | 112 | 12 |
| October.......... | 117 | 100 | 100 | 103 | 97 | 100 | 109 | 102 | 111 | 11 |
| November........ | 102 | 98 | 89 | 100 | 98 | 98 | 107 | 100 | 108 | 8 |
| December........ | 104 | 79 | 112 | 100 | 102 | 102 | 109 | 100 | 109 | 9 |

movements and of chance influences are minimized. The secular trend is satisfactorily eliminated by the process of adjusting. The index thus obtained should be accurate for the years just preceding the establishment of the Federal Reserve System. Since that time seasonal changes have lessened.[1]

**Business Cycles.** The *Review of Economic Statistics,* in the analyses just referred to, has made exhaustive studies of the cyclic movements of the commonly used business barometers. After measuring the influence of seasonal variations by a process similar to that just described, it determined the cycles on the basis of lines of least squares. It then combined twelve of the principal barometric indexes into three composite indexes which are taken as measurements of speculative activity, business activity, and banking strain, respectively. A chart of these three composite indexes for the years 1903 to 1913, inclusive, is here reproduced (Figure 13). The chart shows very clearly the general stages of the business cycle, from the predominance of the speculative activity which marks the awakening from a period of depression, through the period of intensified production, and into the period of banking strain which heralds another depression. The various barometric series used in constructing the figure are indicated in the explanation accompanying the title. It should be added that while New York bank loans point to the speculative aspects of the cycle, those outside New York conform more closely to business

[1] For the most exhaustive study of seasonal variations in the interest rate, the student should consult ''Seasonal Variations in the Relative Demand for Money and Capital in the United States'' (National Monetary Commission, 1910), by E. W. Kemmerer. See also the *Monthly Review,* Federal Reserve Bank of New York, Feb. 1, 1922.

activity. For practical purposes measurements of the general aspects of the cycle need to be supplemented by indexes showing the position of particular industries, as indicated by relative production and stocks of goods on hand. This need is now being met in part by the Federal Reserve Bank of New York, and by other agencies.

Several phenomena more psychological than economic in nature show a tendency to fluctuate more or
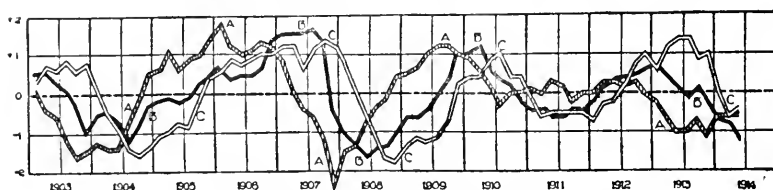


FIGURE 13. The index of general business conditions, 1903-14.

A, Speculation: New York Bank clearings, average price of industrial stocks, average price of railroad stocks, and average price of railroad bonds.

B, Business: Bank clearings of the United States outside New York City, Bradstreet's index of wholesale commodity prices, United States Bureau of Labor Statistics' index of wholesale commodity prices, and pig-iron production.

C, Banking: Interest rates on 60-90 day and on 4-6 months commercial paper in New York City, loans and deposits of New York City clearing house banks (both inverted).

Reproduced from *Review of Economic Statistics,* by permission of the editors.

less closely with the business cycle. Professor A. H. Hansen has recently shown statistically that since 1898 strikes increased with prosperity, though before that date, when the trend of prices was downward, they increased with depression (cf. *American Economic Review,* Dec., 1921, pp. 617-621). Mr. Roger Babson has traced a connection between church growth and the business cycle; religious activity being apparently intensified during a period of depression.

Unemployment, failures, suicides, and crime generally, also increase during a period of depression. On the other hand, immigration, the marriage rate, and extravagance, tend to increase with a period of prosperity (cf. Figures 10 and 11, pages 114 and 115).

As an example of the application of statistical methods to the practical problem of forecasting the cyclic movements of business, the "Annalist Barometer and Business Index Line" may be cited. This index is published in graphic form, together with a brief explanation, each week in the *Annalist,* a well-known financial paper of New York. It is derived from the data elaborated by the *Review of Economic Statistics,* already briefly described. The index is the reciprocal of a weighted average of the deviations from normal of commodity prices, interest rates, pig iron production, New York bank clearings, and bank clearings outside of New York. Since these series of data, when directly combined, measure the later phases of the business cycle, their decline precedes and forecasts the rise in stocks marking the beginning of the next cycle; and their rise similarly forecasts a decline in stocks. By taking the reciprocal of the deviations of the combined series, the forecast is made direct instead of inverse. By a comparison of the forecasting index and the movement of stocks in former years, the decisiveness of change in the former necessary to constitute a forecast of the latter has been determined. A detailed account of the construction and use of the index will be found in the *Annalist* of March 28 and of October 24, 1921.

The student who desires to inquire more intensively

into the statistics of the business cycle should consult for himself the data published in the *Review of Economic Statistics* already mentioned. In addition he should become familiar with Wesley C. Mitchell's standard work on *Business Cycles,* and with a more recent work by D. F. Jordan on *Business Forecasting.* In connection with the underlying causes of the cycle, reference should be made to the interesting but very technical works of H. L. Moore (cf. *Economic Cycles,* and articles in the *Quarterly Journal of Economics,* February, August, and November, 1921). Professor Moore discovers some relation to exist between a weather cycle of heavier and lighter rainfall and the business cycle, the average duration of each cycle being about eight years. The moist years bring as a rule larger crops, with some tendency to a lowering of the general price level, followed during the drier years by a rise in the price level. This relation seems to be clearer for English prices than for American. The weather cycle is shown to be synchronous in several countries, and to be correlated with a cycle of barometric pressure, which in turn may have astronomical causes. But while the subject is very interesting and valuable theoretically, the correlations disclosed are too irregular to be of great practical value.

### REFERENCES

Babson, Roger W., *Business Barometers.*

Davies, G. R., "Social Aspects of the Business Cycle," *Quarterly Journal of the University of North Dakota,* January, 1922.

Hurlin, Ralph G., "The Long-Time Trend of Prices in the United States," *The Annalist,* July 4, 1921.

Jordan, D. F., *Business Forecasting.*

Kemmerer, E. W., *High Prices and Deflation.*

Mitchell, Wesley C., *Business Cycles.*

Moore, Henry L., *Economic Cycles: Their Law and Cause.*

Peddle, John B., *The Construction of Graphical Charts,* Chapter VI.

Persons, W. W., "Construction of a Business Barometer Based upon Annual Data," *American Economic Review,* December, 1916, pp. 739-769.

Piatt, Andrew A., *National Monetary Commission.*

Tingley, Richard H., "Another Yardstick of Banking Conditions," *The Annalist,* November 28, 1921, p. 511.

## EXERCISES

1. Plot the data for production and price of wheat, 1870-1920 (Tables XIV and XIV-A, pp. 76 and 77), and draw a free-hand trend for each series.

2. As in Exercise 1, construct free-hand trends for the production and price of corn.

3. Plot the index of physical production of crops, 1870-1920 (Table XV, p. 81), and draw a straight-line trend by inspection.

4. Apply the method of semi-averages to the construction of a straight-line trend for the data of the preceding exercise.

5. Compute a five-year moving average of per capita production, 1890-1918. Plot both the trend and the data from which it is derived on 17"x22" cross-section paper, and measure graphically the deviations from the trend.

6. Plot on a horizontal axis the deviations obtained in the preceding exercise. Find the average deviation, and indicate this on the graph for both the plus and the minus deviations.

7. Compute and plot a straight-line trend (line of least squares) for the following price index.

| Year | Prices (3 articles) |
|------|---------------------|
| 1897 | 85 |
| 1898 | 70 |
| 1899 | 90 |
| 1900 | 130 |
| 1901 | 125 |

8. Find straight-line trends (lines of least squares) for the production and price of pig iron as given below, taking each year separately.

PRODUCTION AND PRICE OF PIG IRON    (*Iron Age*)

(000 omitted from production)

| Month | 1909 | | 1910 | | 1911 | | 1912 | | 1913 | |
|-------|------|-------|------|-------|------|-------|------|-------|------|-------|
| | Tons | $ | Tons | $ | Tons | $ | Tons | $ | Tons | $ |
| Jan.. | 1,797 | 16.25 | 2,608 | 17.25 | 1,759 | 14.25 | 2,057 | 13.25 | 2,795 | 16.95 |
| Feb.. | 1,707 | 16.13 | 2,397 | 17.06 | 1,794 | 14.25 | 2,100 | 13.31 | 2,586 | 16.69 |
| March | 1,832 | 15.05 | 2,617 | 16.30 | 2,188 | 14.25 | 2,405 | 13.50 | 2,763 | 16.31 |
| April. | 1,738 | 14.25 | 2,483 | 15.37 | 2,065 | 14.25 | 2,375 | 13.75 | 2,752 | 15.65 |
| May.. | 1,883 | 14.50 | 2,390 | 15.00 | 1,893 | 13.95 | 2,512 | 14.15 | 2,822 | 14.94 |
| June . | 1,930 | 14.70 | 2,265 | 14.85 | 1,787 | 13.44 | 2,440 | 14.25 | 2,628 | 14.06 |
| July.. | 2,103 | 15.75 | 2,148 | 14.75 | 1,793 | 13.25 | 2,410 | 14.70 | 2,560 | 13.75 |
| Aug.. | 2,248 | 16.38 | 2,106 | 14.31 | 1,926 | 13.45 | 2,512 | 15.06 | 2,543 | 14.06 |
| Sept.. | 2,385 | 17.35 | 2,056 | 14.25 | 1,977 | 13.31 | 2,463 | 15.87 | 2,505 | 14.25 |
| Oct.. | 2,599 | 17.88 | 2,093 | 14.25 | 2,102 | 13.25 | 2,689 | 16.80 | 2,546 | 14.35 |
| Nov.. | 2,547 | 17.75 | 1,909 | 14.25 | 1,999 | 13.20 | 2,630 | 17.25 | 2,233 | 13.87 |
| Dec.. | 2,635 | 17.45 | 1,777 | 14.25 | 2,043 | 13.19 | 2,782 | 17.25 | 1,983 | 13.95 |
| Total | 25,410 | 16.12 | 26,855 | 15.16 | 23,329 | 13.67 | 29,383 | 14.93 | 30,722 | 14.90 |

(Reprinted, with permission, from Babson's Desk Sheet.)

9. Find the average index of per capita physical production in the United States (page 81) for each decade from 1870 to 1919. Using the resulting five averages, construct a line of least squares. Plot the original data and the trend thus found.
10. Find the average index of production of iron and copper by five year periods from 1870 to 1914. Plot these averages, and fit to them a parabola of the second degree.
11. Compute and plot the cycles of the interest rate, 1909-1913, using the data and index of seasonal variations presented in Table XVII, page 117.
12. Using the data given below, find an index of seasonal variations in exports of merchandise (a) by the method of averages, and (b) by the link-relative method.

EXPORTS OF MERCHANDISE, UNITED STATES, 1909-1913

(In Millions of Dollars)

|  | 1909 | 1910 | 1911 | 1912 | 1913 |
|---|---|---|---|---|---|
| January .......... | 157 | 144 | 197 | 202 | 227 |
| February ......... | 126 | 125 | 176 | 199 | 194 |
| March ............ | 139 | 144 | 162 | 205 | 187 |
| April ............. | 125 | 133 | 158 | 179 | 200 |
| May .............. | 123 | 131 | 153 | 175 | 195 |
| June ............. | 117 | 128 | 142 | 138 | 163 |
| July ............. | 109 | 115 | 128 | 149 | 161 |
| August ........... | 110 | 135 | 144 | 168 | 188 |
| September ........ | 154 | 169 | 196 | 200 | 218 |
| October ........... | 201 | 208 | 210 | 255 | 272 |
| November ......... | 194 | 207 | 202 | 278 | 246 |
| December ......... | 172 | 229 | 225 | 250 | 233 |

(December, 1908, 170)

13. The following table, adapted from the Yearbook of the Department of Agriculture, 1918, gives the farm price of wheat in the United States (cents per bushel) on the first day of each month for the years 1909 to 1913 inclusive. Using the method of link relatives, derive an index of seasonal variations.

|  | 1909 | 1910 | 1911 | 1912 | 1913 |
|---|---|---|---|---|---|
| January 1 .... | 93.5 | 103.4 | 88.6 | 88.0 | 76.2 |
| February 1 ... | 95.2 | 105.0 | 89.8 | 90.4 | 79.9 |
| March 1 ...... | 103.9 | 105.1 | 85.4 | 90.7 | 80.6 |
| April 1 ....... | 107.0 | 104.5 | 83.8 | 92.5 | 79.1 |
| May 1 ........ | 115.9 | 99.9 | 84.6 | 99.7 | 80.9 |
| June 1 ....... | 123.5 | 97.6 | 86.3 | 102.8 | 82.7 |
| July 1 ....... | 120.8 | 95.3 | 84.3 | 99.0 | 81.4 |
| August 1 ..... | 107.1 | 98.9 | 82.7 | 89.7 | 77.1 |
| September 1 .. | 95.2 | 95.8 | 84.8 | 85.8 | 77.1 |
| October 1 ..... | 94.6 | 93.7 | 88.4 | 83.4 | 77.9 |
| November 1 ... | 99.9 | 90.5 | 91.5 | 83.8 | 77.0 |
| December 1 ... | 98.6 | 88.3 | 87.4 | 76.0 | 79.9 |

(December 1, 1908, 92.2)

14. Using the method of averages, derive an index of seasonal variations from the following table of farm prices of wheat in the United States (cents per bushel) for the years 1909 to 1918, inclusive.

| Yearly averages | | Monthly averages | |
|---|---|---|---|
| 1909 | 101.3 | Jan. 1 | 109.4 |
| 1910 | 96.5 | Feb. 1 | 115.2 |
| 1911 | 86.9 | March 1 | 115.2 |
| 1912 | 87.4 | April 1 | 116.4 |
| 1913 | 78.4 | May 1 | 125.6 |
| 1914 | 88.4 | June 1 | 126.0 |
| 1915 | 105.2 | July 1 | 117.7 |
| 1916 | 125.9 | Aug. 1 | 117.9 |
| 1917 | 200.8 | Sept. 1 | 117.4 |
| 1918 | 204.3 | Oct. 1 | 116.5 |
| | | Nov. 1 | 119.7 |
| | | Dec. 1 | 118.6 |

15. Compute and plot the cycles in the price of wheat, 1909-1913, as measured from a line of least squares. Use the data of Exercise 13.
16. From financial journals and other sources obtain monthly or weekly quotations for recent and current dates illustrating barometric subjects such as are mentioned on pages 66 and 124. Plot the data and construct trends. On the basis of these barometers and such other information as may be available, make a forecast of business conditions for the immediate future, allowing for seasonal variations.

# CHAPTER VI

## CORRELATION

**Correlation Defined.** A study of the cycles of business barometers leads to the problem of classifying and measuring the relationships among them. These relationships may be discovered in various forms and degrees. For example, the cycles of building permits and pig iron production will be found to move somewhat closely together. Then again, stock prices and commodity prices form similar waves, though the latter usually follow a few months behind. On the other hand, commodity prices and business failures show opposite movements—when one is up the other is down. All such relationships between two sets of data are known as correlations. When the two sets of cycles agree, the correlation is called positive; when they disagree, it is negative. When the cycles are not quite coincident in point of time, the one which follows is said to show a "lag" of a given interval.[1]   Correla-

---

[1] The term "lag" is also sometimes used to designate a smaller degree of variation occurring in one series than in another comparable to it. Thus the lag of retail prices behind wholesale prices is largely a matter of degree, and only slightly a matter of time. But it is the time element only that enters into the calculation of correlation. In allowing for the lag, the series coming later in time is considered as moved back by the length of the lag, and the corresponding items are then compared. When the length of the lag is difficult to determine, estimates must be made, and the correlation computed on the basis of each estimate. The lag resulting in the most marked correlation is assumed to be the correct one. In determining the lag it is often necessary to take into account the causal relation existing between the two series under consideration, as in a case where the assumption of a lag for one series results in a positive correlation, while the transfer of the lag to the other series results in a negative correlation.

tion carries the idea of a fundamental relationship: either one phenomenon acts or reacts upon the other, or both are due to common causes. The principle is not limited to time series. Comparison might be made, for example, between the advertising and the rate of earnings of given business firms. But in any case the principle would be the same as before, and the methods used would be practically identical.

**The Graphic Method.** A fairly good study of correlation in economic phenomena can often be made without any more elaborate methods than those already described in isolating and plotting the cycles. Two time series, reduced to standard deviation cycles and plotted on equal horizontal scales may be very well compared by superimposing one on the other. To facilitate comparison, one may be drawn on a transparent medium, such as tracing cloth; or a mimeoscope may be used. By shifting the superimposed cycles back and forth, the lag may be fairly accurately determined. The correlation may be described as positive or negative, high, moderate or low, and the lag and its consistency may be stated. This is the method adopted by the *Review of Economic Statistics* in its study of the correlations existing among 24 business barometric series for the years 1903-1914.

**Method of Concurrent Deviations.** It is often desirable, however, to measure the degree of correlation in precise mathematical terms. This is particularly true when correlated data are being used in support of a given theory. In order to obtain a precise result, mathematical methods developed originally for use in biometrics have been borrowed and adapted.

As an introduction to the mathematical methods of measuring correlation, we may take up a simple formula which is well adapted to the comparison of short-time fluctuations. The formula is the expression of what is called the method of concurrent deviations. It may be illustrated by applying it to a comparison between the short-time fluctuations of real wages and per capita production in the United States (cf. pp. 51, 53, and 81). The fluctuations may be most readily determined by reference to a graph of each series (cf. Fig. 6, p. 101). If at any given year the line makes an inverted angle, like a caret ($\Lambda$), the fluctuation is registered on the index as positive ($+$). If the angle is V-shaped, it is registered as negative ($-$). If no angle is formed, the year is indicated as neutral (0). In some cases it may not be possible to determine from the graph whether the angle is neutral, or slightly positive or negative; in which case resort may be had to the data.

After the deviations of both series have all been registered, they are compared across, item by item. If in a given year both indexes show a positive fluctuation, one agreement is counted. If positive and negative meet, one disagreement is counted. If one or both of the fluctuations of a given year are neutral, one-half is added to both the agreements and disagreements. When this summation is complete, the larger of the two totals thus obtained is designated as the number of concurrent deviations, denoted in the formula by the letter C. The sign of the coefficient to be obtained by the use of the formula is determined by the nature of the concurrences. If they are agree-

ments, the sign is positive; if disagreements, the sign is negative. The formula for correlation (R) as thus measured is:

$$R = \pm \sqrt{\frac{2C-N}{N}}$$

In the case of the wage and production indexes just mentioned, the number of disagreements, or concurrences, totals to $33\frac{1}{2}$ and the number of comparisons is 49. The formula therefore becomes:

$$R = -\sqrt{\frac{67-49}{49}} = -.61$$

The derivation of the formula is of little importance, as it is patterned empirically on the one next to be described. The significance of the coefficient will become evident in the same connection.

**The Pearson Method.** We come now to the so-called Pearson "r", the most satisfactory method to apply to straight line correlations: that is, to those which when graphed show an approximation to a straight line rather than a curve. This statement does not refer to the trends of the two series taken separately, but only to the trend formed by the two sets of cycles plotted as $x$ and $y$, respectively. In explaining the method, it is most convenient to begin with two sets of cycles, or deviations, already reduced to units consisting of their respective standard deviations. The following table gives two such series, and the process of finding their correlation. The two cycles are graphed as coordinates in Figure 14, page 136.

Since the units used in both cases are standard deviations, the spread on the two axes, as measured by the

CORRELATION OF PRICES (X) AND EMPLOYMENT (Y)
JANUARY, 1920, TO JANUARY, 1921

Cycles, in Units of Standard Deviation

| x | y | $x^2$ | xy |
|---|---|---|---|
| .75 | .59 | .5625 | .4425 |
| .94 | .85 | .8836 | .7990 |
| .91 | .51 | .8281 | .4641 |
| .88 | 1.10 | .7744 | .9680 |
| .89 | .68 | .7921 | .6052 |
| .57 | .51 | .3249 | .2907 |
| .37 | .34 | .1369 | .1258 |
| .18 | .42 | .0324 | .0756 |
| −.14 | −.09 | .0196 | .0126 |
| −.53 | −.34 | .2809 | .1802 |
| −.98 | −.59 | .9604 | .5782 |
| −1.74 | −1.27 | 3.0276 | 2.2098 |
| −2.10 | −2.71 | 4.4100 | 5.6910 |

Totals    0             0          13.0334      12.4427

$$r = \frac{12.4427}{13.0334} = .96$$

cycles squared, must necessarily be equal. If every
deviation in one series concurs with an equal deviation
in the other series, the points when plotted will neces-
sarily fall on a diagonal sloping upward from left to
right at 45°. If positive deviations concur with nega-
tive, the points will lie in a diagonal sloping downward
from left to right at 45°. In the first case a line of least
squares drawn through the points will necessarily have
a slope of + 1, and in the second case of −1. These
are obviously the largest results, both positive, and
negative, that could be obtained from two such correl-
ative series. A neutral result of zero would be ob-
tained if the points as plotted fall in haphazard posi-
tions about the two axes. The slope of the line of
least squares (the tangent of its angle with the X-axis)

is therefore taken as the measure of correlation. Its basic formula is:

$$r = \frac{\Sigma \, x \, y}{\Sigma \, x^2}$$

In ordinary work it is, of course, necessary first to find a trend for each series, if the cycles are to be measured. If the deviations are taken from the average of each series, the general direction and form of the two lines will be contrasted. This is equivalent to assum-
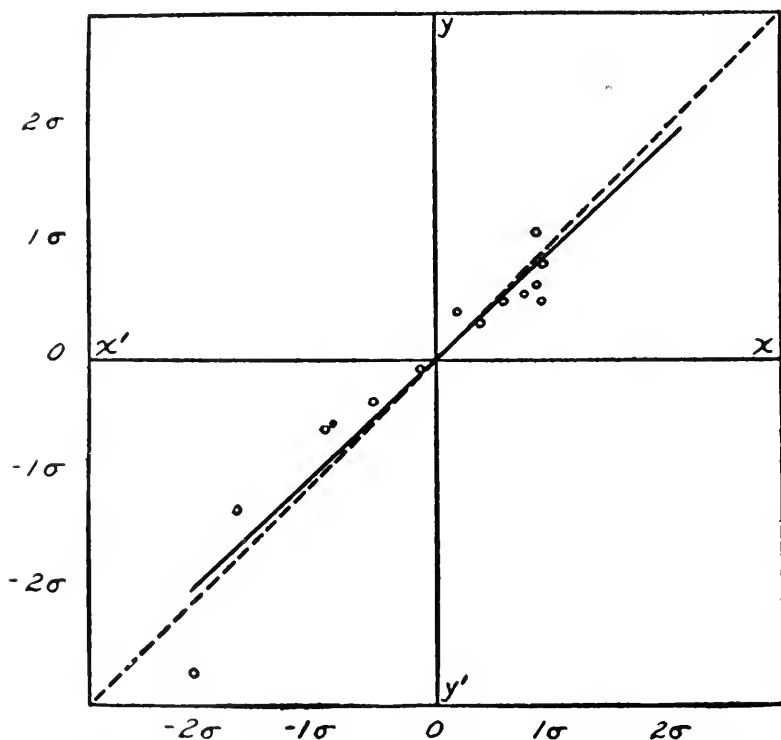


FIGURE 14. Correlation of price cycles (x) and employment cycles (y), expressed in units of the standard deviation of each series, respectively. Line of least squares (solid line), and line of unit slope (broken line).

ing a horizontal trend as a basis of measurement. In certain cases this comparison may be desired. But usually it is advisable to measure either the two trends by eliminating the cycles or to measure the cycles as taken from the trend. The latter is the usual procedure, since interest generally lies in a comparison of the cycles.

The computing of a coefficient of correlation from data which require the finding of trends is illustrated in Table XVIII. The work is for the most part self-explanatory, since the trends are found by processes already explained. Instead, however, of using the standard deviations as the units in which to express the cycles, the original units are employed throughout. The reduction to standard deviation units is in effect obtained by inserting $\sigma_1$ and $\sigma_2$ in the denominator of the formula for $r$. But a substitute must be found for $\Sigma x^2$, which in the formula as just used was also in terms of the standard deviation. The required substitute is N. This may be seen by recalling that in the standard deviation series $\sqrt{\dfrac{\Sigma x^2}{N}} = \sigma = 1$. Hence $\Sigma x^2$ must equal N. The formula for the line of least squares as applied to the previous problem in correlation may therefore be transformed into the Pearson correlation formula, thus:

$$r = \frac{\Sigma xy}{\Sigma x^2}$$ (both $x$ and $y$ being expressed in units of their respective standard deviations.)

$$= \frac{\Sigma xy}{N \sigma_1 \sigma_2}$$ (in which $x$ and $y$ are expressed in terms of the original units.)

$$= \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$ (an alternate form obtained by substitution.)

TABLE XVIII

CORRELATION OF ANNUAL PRODUCTION OF PIG IRON AND AVERAGE PRICE, U. S., 1903-1913

(Cycles measured from Line of Least Squares *)

| YEAR | PRODUCTION (MILLIONS OF TONS) | x | xy | TREND | PRICE PER TON | x | xy | TREND |
|---|---|---|---|---|---|---|---|---|
| 1903 | 18.0 | -5 | -90.0 | 18.2 | $17.10 | -5 | -85.5 | 16.8 |
| 1904 | 16.5 | -4 | -66.0 | 19.3 | 12.70 | -4 | -50.8 | 16.6 |
| 1905 | 23.0 | -3 | -69.0 | 20.4 | 15.60 | -3 | -46.8 | 16.4 |
| 1906 | 25.3 | -2 | -50.6 | 21.6 | 16.70 | -2 | -33.4 | 16.3 |
| 1907 | 25.8 | -1 | -25.8 | 22.7 | 23.10 | -1 | -23.1 | 16.1 |
| 1908 | 15.9 | 0 | 0 | 23.8 | 15.50 | 0 | 0 | 16.0 |
| 1909 | 25.8 | 1 | 25.8 | 25.0 | 16.10 | 1 | 16.1 | 15.8 |
| 1910 | 27.3 | 2 | 54.6 | 26.1 | 15.20 | 2 | 30.4 | 15.6 |
| 1911 | 23.7 | 3 | 71.1 | 27.2 | 13.70 | 3 | 41.1 | 15.5 |
| 1912 | 29.7 | 4 | 118.8 | 28.3 | 14.90 | 4 | 59.6 | 15.3 |
| 1913 | 31.0 | 5 | 155.0 | 29.5 | 14.90 | 5 | 74.5 | 15.2 |
| | 262.0 | $\Sigma x^2 = 110$ ) 123.9 | | | 175.50 | $\Sigma x^2 = 110$ ) -17.9 | | |
| | A = 23.8 | S = 1.13 | | | A = 16.0 | S = .16 | | |

* If the eyes have been found by a method which fails appreciably to center them, it is not necessary to work out an adjustment. In finding the standard deviation, the usual correction for an inexact average may be made. The $\Sigma xy$ may also be corrected as shown in Table XIX. The student should be careful to distinguish the product $xy$ obtained in finding the trend from the corresponding product obtained in finding the coefficient of correlation. For convenience the latter is here designated by small capitals.

| YEAR | PRODUCTION CYCLES X | X² | PRICE CYCLES Y | Y² | XY |
|---|---|---|---|---|---|
| 1903 | -.2 | .04 | .3 | .09 | -.06 |
| 1904 | -2.8 | 7.84 | -3.9 | 15.21 | 10.92 |
| 1905 | 2.6 | 6.76 | -.8 | .64 | -2.08 |
| 1906 | 3.7 | 13.69 | .4 | .16 | 1.48 |
| 1907 | 3.1 | 9.61 | 7.0 | 49.00 | 21.70 |
| 1908 | -7.9 | 62.41 | -.5 | .25 | 3.95 |
| 1909 | .8 | .64 | .3 | .09 | .24 |
| 1910 | 1.2 | 1.44 | -.4 | .16 | -.48 |
| 1911 | -3.5 | 12.25 | -1.8 | 3.24 | 6.30 |
| 1912 | 1.4 | 1.96 | -.4 | .16 | -.56 |
| 1913 | 1.5 | 2.25 | -.3 | .09 | -.45 |
|  | 14.3 | 11)118.89 | 8.0 | 11)69.09 | 40.96 |
|  | -14.4 | σ²=10.808 | -8.1 | σ²=6.281 |  |
|  |  | σ=3.29 |  | σ=2.51 |  |

$$r = \frac{\Sigma xy}{N\sigma_1\sigma_2} = \frac{40.96}{11 \times 3.29 \times 2.51} = .45$$

$$\text{P.E.} = \frac{.6745(1-r^2)}{\sqrt{N}} = \frac{.6745(1-.20)}{3.32} = .16$$

**The Probable Error.** It will be seen that the first part of Table XVIII records merely the finding of the trends (lines of least squares) for the two series. The figures have not been carried out beyond approximately one per cent accuracy. The production and price cycles shown in the latter part of the table are obtained by the usual method of subtracting the trend items
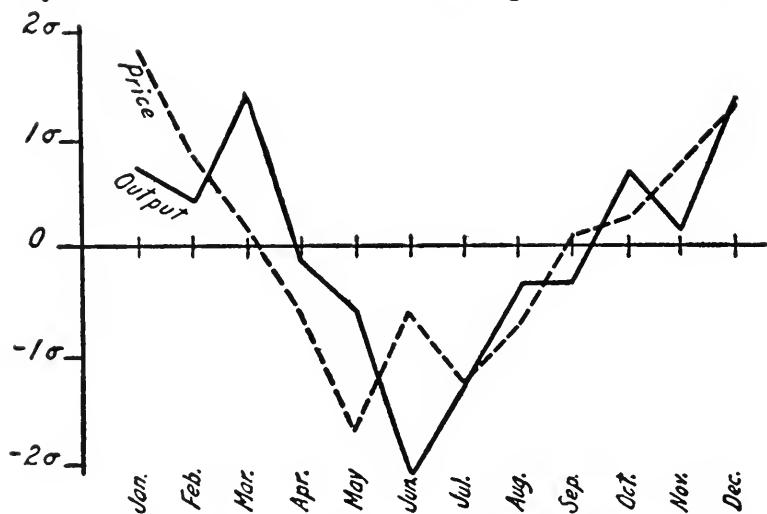


FIGURE 15. Variations in production and price of pig iron, United States, 1908. Standard deviation units, measured from the line of least squares (r = .68).

from the corresponding items in the original series. The standard deviation of each of the resulting series of cycles and the summation of the xy's are next found. The Pearson formula is then applied, and the result, r = .45, is obtained. To this result is appended what is known as the "probable error" (P.E.), the word "error" being here used merely in the sense of divergence, as in the theory of least squares. The formula for the probable error expresses the quartile

deviation from the coefficient of .45 which would normally appear by the operation of the laws of chance. The probable error therefore gives some idea of the range over which the value of $r$ has an even chance of deviating, and may be used in estimating the significance of the correlation. On the basis of experience it is assumed that if the value of $r$ is as low as thirty, or if the probable error is as high as one-third of $r$,
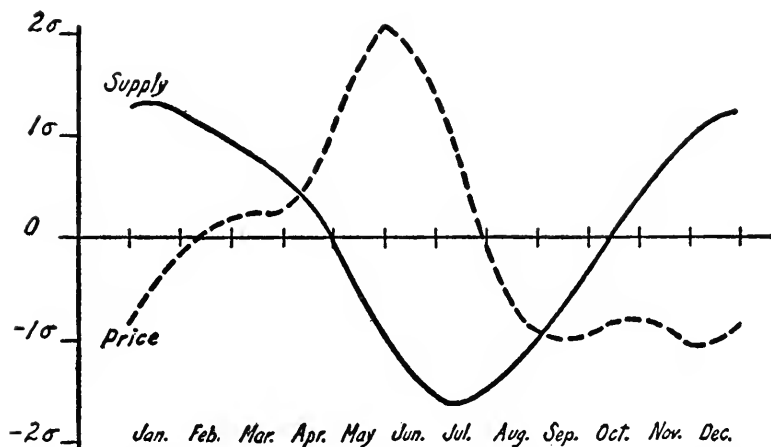


FIGURE 16. Seasonal variations in the visible supply and the price of wheat, United States, 1909-1913 ($r = -.87$, prices preceding one month). For data, see Exercise 8, page 150.

correlation is barely indicated. If, however, $r$ is as high as fifty, and if the probable error is not more than one-fifth of $r$, correlation is clearly indicated. Between these limits a correlation may be regarded as more or less tentatively indicated.

**The Relation of Output to Price.** The result obtained in Table XVIII indicates that in the iron industry production is directly adjusted to meet demand as reflected in the price. When the price is high, pro-

duction therefore is high, and vice versa. This relation is seen to be very close in certain years if monthly data are used (cf. Figure 15). The case is doubtless typical of a large part of manufacturing, particularly when the process is relatively short. But in agriculture, where the maturing of the output is determined by the seasons, contemporaneous movements of output and prices are negatively related. This fact is shown indirectly by Figure 16. About 75% of the world's wheat crop is harvested in the months of June, July, and August, with a consequent rapid depression
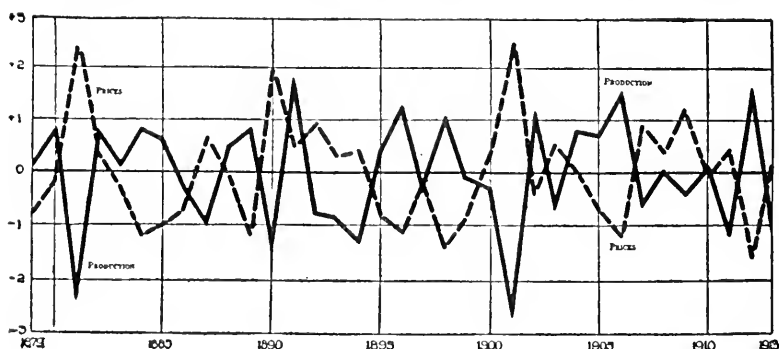


FIGURE 17. Comparison of corrected figures for index of physical production of crops and index of crop prices. The long time movements, or secular trends, have been eliminated, and the two series have been expressed in comparable units.
Reproduced from the *Review of Economic Statistics,* by permission of the editors.

of the price. Extremely large crops are normally followed for several months by unusually low prices, and vice versa. This fact is depicted in Figure 17, where agricultural production is plotted against the subsequent price, rather than the average price for the year. A comparison of the cycles of crops and of manufactures suggests that the former exert some pos-

## TABLE XIX
### CORRELATION OF ENTRANCE GRADES AND SCHOLARSHIP GRADES

| D | Entrance grade | \ | SCHOLARSHIP GRADES—AVERAGE FOR FOUR YEARS | | | | | | | | | | | | | | F | FD | FD² |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 4¾ | 4½ | 4¼ | 4 | 3¾ | 3½ | 3¼ | 3 | 2¾ | 2½ | 2¼ | 2 | 1¾ | 1½ | | | |
| | | | -6 | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | | |
| 4 | 90 | | | | | | | | | | | | | | 1 | | 1 | 4 | 16 |
| 3 | 85 | | | | | | | | | | | | 1 | 2 | 1 | | 4 | 12 | 36 |
| 2 | 80 | | | | | | | | | | 2 | 3 | 2 | | | 7 | 14 | 28 |
| 1 | 75 | | | | | | | | 3 | 3 | 2 | | 2 | | | | 10 | 10 | 10 |
| 0 | 70 | | | | | 4 | 4 | 3 | 6 | | | | | | | | 17 | 0 | 0 |
| -1 | 65 | | 1 | 2 | 2 | 3 | 2 | | | | | | | | | | 10 | -10 | 10 |
| -2 | 60 | | 1 | | 1 | 1 | | | | | | | | | | | 3 | -6 | 12 |
| | **F =** | | 2 | 2 | 3 | 8 | 6 | 3 | 9 | 3 | 2 | 2 | 6 | 4 | 1 | 1 | =52 | )24 | 112 |
| | **FD =** | | -12 | -10 | -12 | -24 | -12 | -3 | 0 | 3 | 4 | 6 | 24 | 20 | 6 | 7 | | | |
| | **FD² =** | | 72 | 50 | 48 | 72 | 24 | 3 | 0 | 3 | 8 | 18 | 96 | 100 | 36 | 49 | | | |

Entrance examination grades—weighted averages.

Right‑hand column summary (entrance grades):

$$K = .461$$
$$K^2 = .213$$
$$\sigma^2 = 1.941$$
$$\sigma = 1.39$$

(with 2.154 and .213)

Bottom‑row summary (scholarship grades):

$$= -3$$
$$K = -.058$$
$$= 579$$
$$(\div N)\ 11.135$$
$$K^2 = .003$$
$$\sigma^2 = 11.132$$
$$\sigma = 3.34$$

$$\Sigma xy = 148$$

$$r = \frac{\Sigma xy - N K_1 K_2}{N \sigma_1 \sigma_2}$$
$$= \left(\frac{\Sigma xy}{N} - K_1 K_2\right)\frac{1}{\sigma_1 \sigma_2}$$
$$= [2.846 - (-.027)]\frac{1}{1.39 \times 3.34} = .62; \quad P.E. = .06$$

itive influence upon the latter within a period of a year or two, but the relation is not very regular.[1]

**Correlation from Frequency Tables.** A somewhat difficult application of the Pearson method of measuring correlation is encountered when the two series which are to be compared are each compiled in frequency tables. The case is illustrated in Table XIX, where average entrance examination grades and average scholarship grades for the four college years are compared. The entrance examination data are expressed in per cents, tabulated to the nearest multiple of five. The scholarship grades were expressed primarily in six groups, ranking downward in order from the first to the sixth. The averaging of such groups for four years gave results carried out to fourths of a group, as shown in the table. For convenience of calculation the values of both scales are converted into unit intervals measuring the deviations, the new scales centering at a zero set opposite the values (70 and $3\frac{1}{4}$) which are assumed as the averages. The number of frequencies for the combined series is written at the appropriate coordinate points in the body of the table, while the frequencies for each scale taken independently are written to the right and below (column F and row F).

The initial steps in the computation will be readily understood by reference to the short-cut method of finding the standard deviation. By this method the standard deviations for both series, respectively, are determined. The finding of $\Sigma xy$ is a somewhat long

---

[1] H. L. Moore, in *Economic Cycles*, finds a positive correlation between an eight-year crop cycle and general prices, allowing a four-year lag to the latter.

process, since each of the frequencies at the coordinate points in the body of the table must be taken into account. Each of these frequencies is multiplied by its two coordinate values, and the products are totaled. To illustrate, the first three columns of frequencies give the following results:

| F | x | y | Fxy |
|---|----|----|-----|
| 1 | –6 | 0 | 0 |
| 1 | –6 | –2 | 12 |
| 2 | –5 | –1 | 10 |
| 1 | –4 | 1 | –4 |
| 2 | –4 | –1 | 8 |

By continuing this computation, a total result of 148, as the value of $\Sigma xy$, is obtained.

Since the two inserted scales measuring the deviations are not centered with precision at the two axes of the table, as determined by the averages of the two series, respectively, the plus and minus deviations from the assumed averages will not exactly balance. Hence a correction must be made in the $\Sigma xy$, just as in the two standard deviations. The corrections ($K_1$ and $K_2$) applied to the finding of the standard deviations are, of course, merely $\Sigma FD \div N$. It may be shown that $\Sigma xy$ will be increased by the product of the two corrections, for every item included. The corrected summation of the moments about the coordinate axes is therefore expressed:

$$\Sigma xy - NK_1K_2$$

In other respects the formula is as previously used. For convenience, however, it is written in the revised form shown at the foot of the table.[1] Applying the

[1] In correlations where the deviations to be contrasted are intended to be taken from the average of each series, the coefficient may be

formula to the data in question, the value of $r$ is found to be .62 $\pm$ .06, a well-marked correlation.

TABLE XX

CORRELATION OF RANKING OF STATES IN MANUFACTURING AND IN LITERACY, U. S., 1860

| STATE | RANK IN MANUFACTURES (CAPITAL PER SQUARE MILE) | RANK IN LITERACY OF NATIVE WHITES | D | D² |
|---|---|---|---|---|
| Alabama | 24 | 23 | 1 | 1 |
| Arkansas | 29 | 24 | 5 | 25 |
| Connecticut | 3 | 2 | 1 | 1 |
| Delaware | 8 | 21 | 13 | 169 |
| Florida | 28 | 22 | 6 | 36 |
| Georgia | 23 | 27 | 4 | 16 |
| Illinois | 16 | 14 | 2 | 4 |
| Indiana | 14 | 17 | 3 | 9 |
| Iowa | 26 | 13 | 13 | 169 |
| Kentucky | 15 | 25 | 10 | 100 |
| Louisiana | 25 | 18 | 7 | 49 |
| Maine | 12 | 5 | 7 | 49 |
| Maryland | 9 | 15 | 6 | 36 |
| Massachusetts | 2 | 1 | 1 | 1 |
| Michigan | 17 | 9 | 8 | 64 |
| Mississippi | 27 | 16 | 11 | 121 |
| Missouri | 19 | 19 | 0 | 0 |
| New Hampshire | 7 | 6 | 1 | 1 |
| New Jersey | 4 | 11 | 7 | 49 |
| New York | 6 | 7 | 1 | 1 |
| North Carolina | 22 | 29 | 7 | 49 |
| Ohio | 10 | 12 | 2 | 4 |
| Pennsylvania | 5 | 10 | 5 | 25 |
| Rhode Island | 1 | 8 | 7 | 49 |
| South Carolina | 21 | 20 | 1 | 1 |
| Tennessee | 18 | 28 | 10 | 100 |
| Vermont | 11 | 3 | 8 | 64 |
| Virginia | 13 | 26 | 13 | 169 |
| Wisconsin | 20 | 4 | 16 | 256 |
| | | | | 1618 |

$$r = 1 - \frac{6\Sigma D^2}{N(N^2-1)} = 1 - \frac{6\times 1618}{29\times 840} = .60$$

P.E. $= .10$

found directly from the original items. This is done by the use of the formula given in Table XIX. An average of zero is assumed for both series, and the items are treated as positive deviations from this average. The standard deviations are found by the modified formula explained on page 41.

**The Method of Rank-differences.** One further modification of the Pearson method of correlation, known as the method of rank-differences, may be noted. This method has the advantage of simplicity, and is especially applicable to comparisons which are made on the basis of approximate measurements only. In Table XX this method is illustrated by applying it to a comparison of the ranking of twenty-nine states in 1860 for manufacturing and literacy. The rankings as here shown are based upon the census of 1860. In arranging such rankings, ties may sometimes occur. In such a case the average rank of the tied items is applied to each of the items. Thus if the second and third items happen to be equal, each is ranked $2\frac{1}{2}$; if the second, third, and fourth are equal, each is ranked 3. When the rankings have been tabulated, as shown, the difference between the two ranks for each state is found. These differences are then squared, and the squares totaled. The formula, as given at the foot of the table, is an adaptation from the one last discussed. Applying the formula, we find that a correlation of .60 exists between the two series. This comparison is an illustration of a number of interesting relationships which may be shown to exist between the economic and the social environment.

**Conclusion.** The purpose of this chapter will have been served if the student has gained a knowledge of the simpler methods commonly employed in measuring correlation. The full theory of the subject is very complex, and is hardly within the scope of an introductory course. A caution must be sounded, however, against an undiscriminating application of the meth-

ods here explained. In particular, conclusions stating causal relationships should never be based on mathematical processes alone. The data, their methods of collection, and the concrete realities they are assumed to measure, must all be subjected to careful scrutiny. The same caution may indeed very properly be extended to the whole field of statistical methods. These methods should prove to be valuable tools in the interpretation of physical, biological, and social phenomena, but they may be a source of positive error if their use is not directed by an adequate comprehension of the field of knowledge in which they are employed.

## *REFERENCES*

Bowley, Arthur L., *Elements of Statistics* (4th Edition), Part II, Chapters VI-IX.

Jevons, W. Stanley, *The Principles of Science.*

King, W. I., ''The Correlation of Historic Economic Variables,'' *Quarterly Publications of the American Statistical Association*, December, 1917, pp. 847-853.

Persons, W. W., ''The Correlation of Economic Statistics,'' *Quarterly Publications of the American Statistical Association*, December, 1910, pp. 287-322.

Secrist, Horace, *Readings and Problems in Statistical Methods*, Chapter X.

West, Carl S., *Introduction to Mathematical Statistics.*

Yule, G. U., *An Introduction to the Theory of Statistics,* Chapters IX-XII.

## EXERCISES

1. On separate sheets of cross-section paper having the same horizontal scale, and with the vertical scales so adjusted as to bring the deviations as nearly as possible to the same measured average, plot the cycles of production and price as obtained in exercises 1 and 2 of the preceding chapter. Similarly plot the cycles in the interest rate, measuring them from Figure 12, and the

price cycles as shown in Figure 6 (pp. 118 and 101). Copy these cycles on tracing paper. Describe the correlation of production and price of the two crops (allowing a lag of one year for prices), and of the interest and price cycles.

2. By the method of concurrent deviations, measure the following correlations (Tables X, XI, and XV, pp. 51, 53, and 81): (a) Wholesale prices and per capita production (both concurrently, and allowing a lag of one year for prices), and (b) Wholesale prices and real wages.

3. Using the table given in exercise 8 of the preceding chapter, and the lines of least squares there obtained, measure by the Pearson ''r'' the correlation of production and price of iron for each year there studied.

4. From the data on page 53, find the correlation between wages and the cost of living for the years 1913-1920 inclusive (Pearson ''r''). Measure the deviations from the average of each series, respectively; that is, assume a horizontal trend.

5. Reduce the deviations obtained in the preceding exercise to units of the standard deviation of each series, respectively, and plot as coordinates the two sets of deviations thus obtained. Compute the line of least squares for the points so plotted, and show that the slope of this line is identical with the coefficient of correlation.

6. Correlate the following indexes (Pearson ''r'') taking the deviations from the average, without finding a trend. Explain the significance of the result.

| Year | Prices | Unemployment |
|---|---|---|
| 1912 | 110 | 70 |
| 1913 | 100 | 120 |
| 1914 | 90 | 140 |
| 1915 | 90 | 100 |
| 1916 | 110 | 70 |

7. Find the Pearson coefficient of correlation for the indexes of normal seasonal variations of merchandise exports from the United States and the price of sterling exchange at New York, as follows:

| Month | Exports | Sterling |
|---|---|---|
| January ..................... | 110 | 100 |
| February ................... | 95 | 108 |
| March ..................... | 99 | 109 |
| April ...................... | 90 | 115 |
| May ....................... | 87 | 116 |
| June ...................... | 80 | 120 |
| July ...................... | 78 | 119 |
| August .................... | 85 | 106 |
| September ................. | 98 | 74 |
| October ................... | 125 | 70 |
| November ................. | 123 | 80 |
| December ................. | 130 | 83 |

8. (a) Find the Pearson coefficient of correlation measuring the relationship between the following indexes of seasonal variation in the visible supply and the price of wheat in the United States, based on the years 1909-1913. (b) Find the coefficient, as before, but assuming that prices tend to anticipate the supply by about a month.

| Month | Visible supply | Price (first of mo.) |
|---|---|---|
| January ................. | 139 | 97 |
| February ................ | 130 | 100 |
| March ................... | 122 | 101 |
| April ................... | 112 | 101 |
| May .................... | 89 | 103 |
| June ................... | 69 | 106 |
| July ................... | 52 | 104 |
| August ................. | 60 | 100 |
| September .............. | 77 | 97 |
| October ................ | 99 | 97 |
| November ............... | 118 | 98 |
| December ............... | 133 | 96 |

9. The following correlation table presents entrance groups (vertical scale) and scholarship groups (horizontal scale) for a certain class of students. Find the coefficient of correlation (Pearson ''r'') and the probable error.

| V | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
|---|----|---|---|---|---|---|---|---|---|---|
| 1 |    |   |   | 2 |   | 1 |   | 2 |   |   |
| 2 |    |   |   |   | 1 | 1 | 1 | 1 |   |   |
| 3 |    | 1 |   | 1 | 4 | 2 |   | 3 | 1 | 1 |
| 4 | 1  | 1 |   | 2 | 1 | 1 | 1 | 1 |   |   |
| 5 | 1  | 3 | 2 |   | 1 |   | 2 |   |   |   |

10. The following correlation table classifies to the nearest twenty-five per cent sixty-nine important commodities according to their price indexes (base, 1913) in May, 1920, and May, 1921 (*Monthly Labor Review*, Aug., 1921, pp. 84-85). The correlation measures approximately the evenness of the price changes occurring between the two dates. Find Pearson's "r."

### Price Indexes, May, 1920

| Price Indexes, May, 1921 | 100 | 125 | 150 | 175 | 200 | 225 | 250 | 275 | 300 | 325 | 350 | 375 | 400 | 425 | 450 | 475 | 500 | 525 | 550 | 575 | 600 | 625 | 650 | 675 | 700 | 725 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 400 |   |   |   |   |   |   |   |   |   | 1 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 1 |
| 375 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 0 |
| 350 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 0 |
| 325 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 0 |
| 300 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 0 |
| 275 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 0 |
| 250 |   |   |   |   |   |   |   |   |   | 1 | 1 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 2 |
| 225 |   |   |   |   |   |   | 1 | 1 | 3 |   | 1 | 1 |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 7 |
| 200 |   |   |   | 1 | 1 | 1 |   |   |   |   | 1 | 1 |   |   | 1 |   |   |   |   |   |   |   |   |   |   |   | 6 |
| 175 |   |   |   | 1 | 1 | 3 |   |   |   |   | 1 | 2 |   | 1 |   |   |   |   |   |   |   |   |   |   |   |   | 9 |
| 150 |   |   |   | 2 |   | 1 | 4 |   |   |   |   |   | 3 |   |   |   |   |   |   | 1 |   |   |   |   |   |   | 11 |
| 125 |   |   | 1 |   | 4 |   | 2 |   |   |   | 1 |   | 2 |   |   | 1 |   | 1 |   |   |   |   |   |   |   |   | 12 |
| 100 | 1 |   | 3 | 4 | 1 |   | 1 | 2 | 1 | 2 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 15 |
| 75 |   | 2 |   |   | 1 |   | 1 | 1 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 1 | 6 |
| **Total** | 1 | 2 | 4 | 8 | 8 | 5 | 9 | 4 | 4 | 4 | 5 | 4 | 5 | 1 | 1 | 1 |   | 1 |   | 1 |   |   |   |   |   | 1 | **69** |

11. The following table shows the ranking of states in (a) Noted men born in state, per 1000 population in 1880, (b) Population per square mile in 1890, and (c) Per cent of urban population in 1890. By the method of rank-differences measure the correlations existing among these three series.

|           | (a) | (b) | (c) |
|-----------|-----|-----|------|
| Alabama ................. | 23 | 24 | 25.5 |
| Arkansas ................. | 29 | 28 | 28 |

| | | | |
|---|---|---|---|
| Connecticut | 4 | 4 | 3 |
| Delaware | 8 | 9 | 11 |
| Florida | 28 | 29 | 20 |
| Georgia | 26 | 22 | 23 |
| Illinois | 16 | 10 | 10 |
| Indiana | 15 | 11 | 17 |
| Iowa | 19 | 20 | 19 |
| Kentucky | 18 | 12 | 21 |
| Louisiana | 25 | 26 | 18 |
| Maine | 5 | 27 | 9 |
| Maryland | 10 | 7 | 8 |
| Massachusetts | 2 | 2 | 2 |
| Michigan | 17 | 18.5 | 14 |
| Mississippi | 27 | 25 | 29 |
| Missouri | 21 | 16 | 16 |
| New Hampshire | 3 | 14 | 6 |
| New Jersey | 11 | 3 | 5 |
| New York | 7 | 5 | 4 |
| North Carolina | 22 | 21 | 27 |
| Ohio | 9 | 8 | 12 |
| Pennsylvania | 12 | 6 | 7 |
| Rhode Island | 6 | 1 | 1 |
| South Carolina | 20 | 17 | 25.5 |
| Tennessee | 24 | 13 | 24 |
| Vermont | 1 | 18.5 | 13 |
| Virginia | 13 | 15 | 22 |
| Wisconsin | 14 | 23 | 15 |

# APPENDIX I

## LABORATORY MATERIAL AND REFERENCES

*Equipment for Graphic Work.* While the larger part of statistical work may be done without any systematic training in mechanical drawing, yet some degree of skill in this field is necessary if graphic representations are to be satisfactorily prepared. The necessary degree of skill may readily be acquired. The student should provide himself with a drawing board, celluloid triangles and irregular curves, a ruler with decimal subdivisions of the inch, a ruling pen, some round-pointed and fine pens for lettering, India ink, several styles of cross-section paper, and a loose-leaf notebook. If he is unfamiliar with the use of drafting materials, he should read the introductory directions given in an elementary treatise on mechanical drawing.

In addition to the material just mentioned, some of the more complicated apparatus used in an engineer's drafting room will be found useful. This equipment may include a drafting machine, a pantagraph, a line-spacer, a map-measurer, and a planimeter. A blue-print outfit is also very useful, indeed is almost a necessity, unless some improved copying device like the photostat is available for use. For elementary work a simple 8″x10″ photography printing frame, and corresponding blue-print paper, will be found quite satisfactory and inexpensive.

*Lettering.* It is not difficult to learn to draw freehand the italic letters used by draftsmen. Directions for such work will be found in "Lettering for Draftsmen, Engineers, and Students," by Charles W. Reinhardt (D. Van Nostrand Company, New York), or in other books on the same subject.

*Other Material.* There are several aids to statistical work that will materially lighten the drudgery incidental to long mathematical processes. The most common of these is the slide rule. A ten-inch rule, giving squares and cubes, will be found sufficient for the greater part of the work involving multiplication, division, and powers or roots. The slide rule

is not difficult to use, and should be mastered by every student of statistics. Besides being an inexpensive and portable device for mathematical operations, it will be found useful in the drawing of logarithmic or ratio graphs, which are now coming into general use. If, however, it is necessary to obtain products or quotients accurate to four or five significant figures, a large cylindrical slide rule may be used, such as the Thatcher, though this is less convenient and considerably more expensive. For powers, roots, and reciprocals, elaborate printed tables are obtainable. If possible, an adding machine (listing) should be available for occasional use, such as the Dalton, the Burroughs, or the Federal. While such a machine is a convenience, a calculating machine is an absolute necessity if very extensive work is to be attempted; and it is well for the student to become acquainted with its operation. Several successful models are now on the market, among which may be mentioned the Burroughs, the Comptometer, the Monroe, and the Marchant. The last two are dial machines, particularly adapted for subtractions and divisions. For certain kinds of statistical work tabulating machines (Hollerith and Powers types) are required, but these machines are so complex and expensive that they can hardly be made available except in the larger laboratories.

*Recording.* Laboratory exercises and other statistical work should be recorded fully, and should be put in clear and neat form. Every graph should be accurately labeled, and the units used in each scale should be indicated. When the scales of a graph do not begin at the zero point, the initial coordinates should not be drawn more heavily than the others, since they are likely to be looked upon as base lines if so drawn. In so far as is practicable, the tables of data from which a graph is drawn should accompany the figure, and the source should be noted. Graphing and lettering should be done in pencil first; the pencil draft may then be completed with India ink, and the pencil lines erased with a soft rubber. Errors in calculation should not be tolerated. All mathematical operations should be performed twice, or some other reliable method of checking should be adopted. Data copied from an original source should always be carefully verified. Tables and mathematical processes may most conveniently be recorded on cross-section paper having one-fifth or one-sixth inch spacing. In any given study the conclusions should be brought out clearly, and their significance explained.

*Various Types of Graphs.* Most of the types of graphs in

common use have been illustrated in the preceding pages. In addition, mention may be made of certain elementary types. One of these is the bar diagram, in which bars of uniform width, and proportional in length to given magnitudes, are used. They are drawn horizontally, except in time series. When they are subdivided, the parts are distinguished by various kinds of cross-hatching and shading. Another is the "pie diagram," in which a circle is subdivided by radii. This diagram is particularly adapted to the representation of percentage subdivisions, such as the relative expenditures for certain classes of goods in a family budget. Another type is the polar chart, designed for graphing seasonal data. Drawings of similar surfaces and solids are sometimes used in the representation of given magnitudes. It should be remembered that, geometrically, magnitudes compared by the use of similar surfaces vary as the square of the dimensions; and by the use of similar solids, as the cube of the dimensions. Sometimes in such drawings it is explicitly stated that the ratio is represented by one dimension only, as when the military forces of different countries are set forth by means of drawings of soldiers whose heights are proportional to the size of the armies. Such drawings are not scientific, however, and are justified only in material of a very popular nature. In general, the use of similar surfaces and solids in the representation of magnitudes is to be avoided. A more complex type of graph is the statistical map. This may be drawn in so many different ways that a general description is impossible. The student having occasion to use it should consult the excellent examples contained in the Statistical Atlas of the United States.

*Sources, References, and Tables.* Brief summaries are given below of the principal sources of statistical material, and the textbook references and statistical tables which are most likely to be of use in connection with an introductory course.

## SOURCES OF STATISTICAL DATA

*Aldrich Report (Senate Report No. 1394)*
*Annalist*
*Bradstreet's*
*Commercial and Financial Chronicle*
*Dun's Review*
*Federal Reserve Bulletin*
*Financial Review (year book)*

*Monthly Labor Review*
*Monthly Review (Federal Reserve Bank of New York)*
*Monthly Summary of Foreign Commerce of the United States*
*Review of Economic Statistics (Harvard)*
*Statesman's Yearbook*
*Statistical Abstract of the United States (yearbook)*
*Statistical services: Babson's, Banker's Statistical Corporation, Brookmire's, and Prentice-Hall.*
*Survey of Current Business.*
*United States Census*
*Weather, Crops, and Markets (U. S. Dept. of Agriculture)*
*World Almanac*
*Yearbook of the Department of Agriculture*

## *TEXTBOOKS, TABLES, AND GENERAL REFERENCES*

*American Economic Review (Bi-monthly)*
Bailey and Cummings, *Statistics*
Barker, E. H., *Computing Tables and Formulas*
*Barlow's Tables*
Bowley, A. L., *Elements of Statistics*
Brinton, W. C., *Graphic Methods for Presenting Facts*
Copeland, M. T., *Business Statistics*
Davenport, C. B., *Statistical Methods*
Jordan, D. F., *Business Forecasting*
*Journal of Political Economy (Monthly)*
King, W. I., *Elements of Statistical Method*
Marshall, Wm. C., *Graphical Methods*
*Quarterly Journal of Economics*
*Quarterly Publications of the American Statistical Association*
Secrist, H., *An Introduction to Statistical Methods*
Secrist, H., *Readings and Problems in Statistical Methods*
West, C. S., *Introduction to Mathematical Statistics*
Whipple, G. C., *Vital Statistics*

# APPENDIX II

## TABLE OF POWERS AND ROOTS

| No. | Square | Cube | Square Root | Cube Root |
|---|---|---|---|---|
| 1 | 1 | 1 | 1.000 | 1.000 |
| 2 | 4 | 8 | 1.414 | 1.259 |
| 3 | 9 | 27 | 1.732 | 1.442 |
| 4 | 16 | 64 | 2.000 | 1.587 |
| 5 | 25 | 125 | 2.236 | 1.709 |
| 6 | 36 | 216 | 2.449 | 1.817 |
| 7 | 49 | 343 | 2.645 | 1.912 |
| 8 | 64 | 512 | 2.828 | 2.000 |
| 9 | 81 | 729 | 3.000 | 2.080 |
| 10 | 100 | 1,000 | 3.162 | 2.154 |
| 11 | 121 | 1,331 | 3.316 | 2.223 |
| 12 | 144 | 1,728 | 3.464 | 2.289 |
| 13 | 169 | 2,197 | 3.605 | 2.351 |
| 14 | 196 | 2,744 | 3.741 | 2.410 |
| 15 | 225 | 3,375 | 3.872 | 2.466 |
| 16 | 256 | 4,096 | 4.000 | 2.519 |
| 17 | 289 | 4,913 | 4.123 | 2.571 |
| 18 | 324 | 5,832 | 4.242 | 2.620 |
| 19 | 361 | 6,859 | 4.358 | 2.668 |
| 20 | 400 | 8,000 | 4.472 | 2.714 |
| 21 | 441 | 9,261 | 4.582 | 2.758 |
| 22 | 484 | 10,648 | 4.690 | 2.802 |
| 23 | 529 | 12,167 | 4.795 | 2.843 |
| 24 | 576 | 13,824 | 4.898 | 2.884 |
| 25 | 625 | 15,625 | 5.000 | 2.924 |
| 26 | 676 | 17,576 | 5.099 | 2.962 |
| 27 | 729 | 19,683 | 5.196 | 3.000 |
| 28 | 784 | 21,952 | 5.291 | 3.036 |
| 29 | 841 | 24,389 | 5.385 | 3.072 |
| 30 | 900 | 27,000 | 5.477 | 3.107 |
| 31 | 961 | 29,791 | 5.567 | 3.141 |
| 32 | 1,024 | 32,768 | 5.656 | 3.174 |
| 33 | 1,089 | 35,937 | 5.744 | 3.207 |
| 34 | 1,156 | 39,304 | 5.830 | 3.239 |
| 35 | 1,225 | 42,875 | 5.916 | 3.271 |

| No. | Square | Cube | Square Root | Cube Root |
|---|---|---|---|---|
| 36 | 1,296 | 46,656 | 6.000 | 3.301 |
| 37 | 1,369 | 50,653 | 6.082 | 3.332 |
| 38 | 1,444 | 54,872 | 6.164 | 3.361 |
| 39 | 1,521 | 59,319 | 6.244 | 3.391 |
| 40 | 1,600 | 64,000 | 6.324 | 3.419 |
| 41 | 1,681 | 68,921 | 6.403 | 3.448 |
| 42 | 1,764 | 74,088 | 6.480 | 3.476 |
| 43 | 1,849 | 79,507 | 6.557 | 3.503 |
| 44 | 1,936 | 85,184 | 6.633 | 3.530 |
| 45 | 2,025 | 91,125 | 6.708 | 3.556 |
| 46 | 2,116 | 97,336 | 6.782 | 3.583 |
| 47 | 2,209 | 103,823 | 6.855 | 3.608 |
| 48 | 2,304 | 110,592 | 6.928 | 3.634 |
| 49 | 2,401 | 117,649 | 7.000 | 3.659 |
| 50 | 2,500 | 125,000 | 7.071 | 3.684 |
| 51 | 2,601 | 132,651 | 7.141 | 3.708 |
| 52 | 2,704 | 140,608 | 7.211 | 3.732 |
| 53 | 2,809 | 148,877 | 7.280 | 3.756 |
| 54 | 2,916 | 157,464 | 7.348 | 3.779 |
| 55 | 3,025 | 166,375 | 7.416 | 3.802 |
| 56 | 3,136 | 175,616 | 7.483 | 3.825 |
| 57 | 3,249 | 185,193 | 7.549 | 3.848 |
| 58 | 3,364 | 195,112 | 7.615 | 3.870 |
| 59 | 3,481 | 205,379 | 7.681 | 3.892 |
| 60 | 3,600 | 216,000 | 7.745 | 3.914 |
| 61 | 3,721 | 226,981 | 7.810 | 3.936 |
| 62 | 3,844 | 238,328 | 7.874 | 3.957 |
| 63 | 3,969 | 250,047 | 7.937 | 3.979 |
| 64 | 4,096 | 262,144 | 8.000 | 4.000 |
| 65 | 4,225 | 274,625 | 8.062 | 4.020 |
| 66 | 4,356 | 287,496 | 8.124 | 4.041 |
| 67 | 4,489 | 300,763 | 8.185 | 4.061 |
| 68 | 4,624 | 314,432 | 8.246 | 4.081 |
| 69 | 4,761 | 328,509 | 8.306 | 4.101 |
| 70 | 4,900 | 343,000 | 8.366 | 4.121 |
| 71 | 5,041 | 357,911 | 8.426 | 4.140 |
| 72 | 5,184 | 373,248 | 8.485 | 4.160 |

| No. | Square | Cube | Square Root | Cube Root |
|-----|--------|------|-------------|-----------|
| 73 | 5,329 | 389,017 | 8.544 | 4.179 |
| 74 | 5,476 | 405,224 | 8.602 | 4.198 |
| 75 | 5,625 | 421,875 | 8.660 | 4.217 |
| 76 | 5,776 | 438,976 | 8.717 | 4.235 |
| 77 | 5,929 | 456,533 | 8.774 | 4.254 |
| 78 | 6,084 | 474,552 | 8.831 | 4.272 |
| 79 | 6,241 | 493,039 | 8.888 | 4.290 |
| 80 | 6,400 | 512,000 | 8.944 | 4.308 |
| 81 | 6,561 | 531,441 | 9.000 | 4.326 |
| 82 | 6,724 | 551,368 | 9.055 | 4.344 |
| 83 | 6,889 | 571,787 | 9.110 | 4.362 |
| 84 | 7,056 | 592,704 | 9.165 | 4.379 |
| 85 | 7,225 | 614,125 | 9.219 | 4.396 |
| 86 | 7,396 | 636,056 | 9.273 | 4.414 |
| 87 | 7,569 | 658,503 | 9.327 | 4.431 |
| 88 | 7,744 | 681,472 | 9.380 | 4.447 |
| 89 | 7,921 | 704,969 | 9.433 | 4.464 |
| 90 | 8,100 | 729,000 | 9.486 | 4.481 |
| 91 | 8,281 | 753,571 | 9.539 | 4.497 |
| 92 | 8,464 | 778,688 | 9.591 | 4.514 |
| 93 | 8,649 | 804,357 | 9.643 | 4.530 |
| 94 | 8,836 | 830,584 | 9.695 | 4.546 |
| 95 | 9,025 | 857,375 | 9.746 | 4.562 |
| 96 | 9,216 | 884,736 | 9.797 | 4.578 |
| 97 | 9,409 | 912,673 | 9.848 | 4.594 |
| 98 | 9,604 | 941,192 | 9.899 | 4.610 |
| 99 | 9,801 | 970,299 | 9.949 | 4.626 |
| 100 | 10,000 | 1,000,000 | 10.000 | 4.641 |

NOTE: In the above table the last two columns are correct to three decimal places, without allowance for decimals dropped.

|  |  | 1850 | 1840 | 1830 | 1800 |  |
|---|---|---|---|---|---|---|
| 1 | 789 | 2,995,772 | 1,793,299 | 1,793,299 | 843,246 | 1 |
| 2 | 521 | 23,191,876 | 17,069,453 | 12,866,020 | 5,308,483 | 2 |
| 3 | 39 | 7.74 | 9.52 | 7.17 | 6.30 | 3 |
| 4 | 000 | 7,135,780,000 | .......... | .......... | .......... | 4 |
| 5 | 93 | 307.69 | .......... | .......... | .......... | 5 |
| 6 | 402 | 63,452,774 | 3,573,344 | 48,565,407 | 82,976,294 | 6 |
| 7 | 91 | 2.74 | 0.21 | 3.77 | 15.63 | 7 |
| 8 | 838 | 63,452,774 | 3,573,344 | 48,565,406 | 82,976,294 | 8 |
| 9 | 587 | 3,752,393 | 174,598 | 1,912,575 | 3,402,601 | 9 |
| 10 | 11 | 0.16 | 0.01 | 0.15 | 0.64 | 10 |
| 11 | 554 | 31,981,739 | 1,675,483 | 643,105 | 317,760 | 11 |
| 12 | 890 | 1,866,100 | 1,726,703 | 2,495,400 | 224,296 | 12 |
| 13 | 775 | 147,395,456 | 79,336,916 | 26,344,295 | 16,000,000 | 13 |
| 14 |  |  |  |  |  | 14 |
| 15 | .. | .......... | .......... | .......... | .......... | 15 |
| 16 | .. | .......... | .......... | .......... | .......... | 16 |
| 17 | .. | .......... | .......... | .......... | .......... | 17 |
| 18 | .. | .......... | .......... | .......... | .......... | 18 |
| 19 | 77 | 131,366,526 | 106,968,572 | 61,000,000 | 10,500,000 | 19 |
| 20 | 452 | 278,761,982 | 186,305,488 | 87,344,295 | 26,500,000 | 20 |
| 21 | 85 | 12.02 | 10.91 | 6.79 | 5.00 | 21 |
| 22 | .. | .......... | .......... | .......... | .......... | 22 |
| 23 | .. | .......... | .......... | .......... | .......... | 23 |
| 24 | 057 | .......... | .......... | .......... | .......... | 24 |
| 25 | .. | .......... | .......... | .......... | .......... | 25 |

# A Picture of the Progress of the United States During 120 Years of National Life

### Compiled from Official Sources by O. P. Austin, Statistician of The National City Bank of New York

*(Reprinted, with permission, from "The Americas")*

| | | 1920 | 1910 | 1900 | 1890 | 1880 | 1870 | 1860 | 1850 | 1840 | 1830 | 1800 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Area — Sq. miles | 3,026,789 | 3,026,789 | 3,026,789 | 3,026,789 | 3,026,789 | 3,026,789 | 3,026,789 | 2,995,772 | 1,793,399 | 1,793,399 | 843,246 | 1 |
| 2 | Population | 105,710,620 | 92,174,415 | 76,303,387 | 62,622,250 | 50,155,783 | 38,558,371 | 31,443,321 | 23,191,876 | 17,069,453 | 12,866,020 | 5,308,483 | 2 |
| 3 | Population, square mile | 34.93 | 30.99 | 25.14 | 20.69 | 16.47 | 12.74 | 10.39 | 7.74 | 9.52 | 7.17 | 6.29 | 3 |
| 4 | Wealth — $ | (*)187,739,071,090 | 107,104,212,000 | 88,517,306,775 | 65,037,091,000 | 42,642,000,000 | 30,068,518,000 | 16,159,616,000 | 7,135,780,000 | . . . . | . . . . | . . . . | 4 |
| 5 | Wealth, per capita — $ | 1,965.00 | 1,161.14 | 1,164.74 | 1,039.25 | 850.20 | 779.83 | 513.93 | 307.60 | . . . . | . . . . | . . . . | 5 |
| 6 | Debt, less cash in Treasury — $ | 24,350,689,711 | 1,046,449,185 | 1,107,711,258 | 890,784,371 | 1,919,326,747 | 2,431,169,956 | 59,964,402 | 63,452,774 | 3,573,344 | 48,565,407 | 82,976,294 | 6 |
| 7 | Debt, per capita — $ | 230.17 | 11.33 | 14.52 | 14.22 | 38.27 | 60.46 | 1.91 | 2.74 | 0.21 | 3.77 | 15.63 | 7 |
| 8 | Interest-bearing debt — $ | 24,230,045,561 | 913,317,499 | 1,023,478,860 | 725,313,110 | 1,723,993,100 | 2,046,455,722 | 64,640,838 | 64,452,774 | 3,573,344 | 48,565,407 | 82,976,294 | 8 |
| 9 | Annual interest charge — $ | 1,016,207,319 | 21,225,902 | 40,919,149 | 29,417,603 | 79,633,981 | 118,784,960 | 3,443,687 | 3,746,390 | 174,598 | 1,912,415 | 4,412,691 | 9 |
| 10 | Interest, per capita — $ | 9.52 | 0.23 | 0.44 | 0.47 | 1.59 | 3.08 | 0.11 | 0.16 | 0.01 | 0.15 | 0.84 | 10 |
| 11 | Gold coined — $ | 16,500,000 | 104,723,735 | 99,272,942 | 20,467,182 | 62,308,279 | 23,198,788 | 23,473,654 | 31,981,789 | 1,675,482 | 643,105 | 317,760 | 11 |
| 12 | Silver coined — $ | 26,957,270 | 3,740,408 | 36,345,321 | 39,202,908 | 27,411,694 | 1,378,256 | 2,259,390 | 1,866,100 | 1,726,703 | 2,495,400 | 224,296 | 12 |
| 13 | Gold in circulation — $ | 879,244,533 | 1,590,572,993 | 610,806,472 | 374,338,928 | 275,605,279 | (*)25,000,000 | . . . . | 147,380,466 | 79,386,916 | 26,144,295 | 16,000,000 | 13 |
| 14 | Silver in circulation — $ | 360,257,853 | 1,296,910,245 | 142,658,344 | 119,711,248 | 68,622,345 | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | 14 |
| 15 | Gold certificates in circulation — $ | 390,665,366 | 802,754,199 | 200,733,019 | 130,830,859 | 7,964,900 | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | 15 |
| 16 | Silver certificates in circulation — $ | 11 — 251,883 | 478,107,258 | 408,406,574 | 297,503,378 | 8,789,500 | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | 16 |
| 17 | U. S. notes in circulation | 337,112,852 | 334,782,870 | 312,971,645 | 324,958,977 | 327,854,437 | 324,962,838 | . . . . | . . . . | . . . . | . . . . | . . . . | 17 |
| 18 | National bank notes in circulation | 696,455,274 | 684,659,585 | 809,113,112 | 147,604,937 | 337,413,778 | 288,648,931 | . . . . | . . . . | . . . . | . . . . | . . . . | 18 |
| 19 | Miscellaneous currency — $ | 1,650,327 | 4,062,525 | 76,909,842 | . . . . | . . . . | 36,602,073 | 207,102,477 | 131,366,028 | 106,968,572 | 61,000,000 | 10,500,000 | 19 |
| 20 | Circulation of money — $ | (*)6,087,256,997 | 3,102,555,995 | 2,066,350,998 | 1,429,251,270 | 973,382,228 | 675,212,794 | 435,407,252 | 278,761,982 | 186,880,488 | 87,444,286 | 26,500,000 | 20 |
| 21 | Per capita — $ | 57.73 | 34.33 | 27.08 | 22.84 | 19.41 | 17.50 | 13.85 | 12.02 | 10.91 | 6.79 | 5.00 | 21 |
| 22 | National banks — No. | 8,030 | 7,145 | 3,732 | 3,484 | 2,076 | 1,612 | . . . . | . . . . | . . . . | . . . . | . . . . | 22 |
| 23 | Capital — $ | 1,224,160,000 | 989,407,114 | 621,536,461 | 642,073,476 | 455,908,903 | 427,233,761 | . . . . | . . . . | . . . . | . . . . | . . . . | 23 |
| 24 | Bank clearings, New York — $ | 252,338,248,490 | 102,533,959,069 | 51,964,588,564 | 37,060,680,572 | 37,182,128,621 | 27,804,539,406 | 7,231,143,667 | . . . . | . . . . | . . . . | . . . . | 24 |
| 25 | Total, United States — $ | 462,920,250,000 | 168,386,563,000 | 84,582,430,051 | 58,844,310,905 | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | 25 |
| 26 | Deposits in national banks — $ | 13,700,150,000 | 5,327,316,312 | 2,458,092,758 | 1,571,714,050 | 831,703,044 | 542,261,583 | 149,277,504 | 43,431,130 | 14,051,520 | 6,914,464 | . . . . | 26 |
| 27 | Savings banks — $ | 6,586,496,000 | 4,070,486,247 | 2,589,719,914 | 1,550,922,958 | 819,106,973 | 549,874,358 | 090,870 | 253,554 | 78,701 | 58,035 | . . . . | 27 |
| 28 | Depositors savings banks — No. | 11,427,556 | 9,142,708 | 6,107,083 | 4,258,893 | 2,335,582 | 1,630,846 | . . . . | . . . . | . . . . | . . . . | . . . . | 28 |
| 29 | Farms and farm property — $ | 77,926,000,000 | 40,991,449,090 | 20,514,001,838 | 16,082,267,689 | 12,180,501,558 | 8,944,857,749 | 7,980,493,063 | 3,967,343,580 | . . . . | . . . . | . . . . | 29 |
| 30 | Farm products, value — $ | 19,858,000,000 | 40,037,000,000 | 5,999,000,000 | 2,460,107,454 | 2,212,540,927 | 1,958,030,927 | . . . . | . . . . | . . . . | . . . . | . . . . | 30 |
| 31 | Manufacturing establishments — No. | 288,376 | 268,491 | 512,254 | 365,455 | 253,852 | 252,148 | 140,433 | 123,025 | . . . . | . . . . | . . . . | 31 |
| 32 | Value of products — $ | 67,888,905,000 | 20,672,051,870 | 13,004,400,143 | 9,372,437,283 | 5,369,579,191 | 4,232,325,442 | 1,885,861,676 | 1,019,106,616 | . . . . | . . . . | . . . . | 32 |
| 33 | Receipts—ordinary — $ | 6,704,453,018 | 675,511,715 | 567,240,852 | 403,080,983 | 333,526,611 | 395,959,833 | 56,054,600 | 43,592,889 | 19,480,115 | 24,844,117 | 10,848,749 | 33 |
| 34 | Customs — $ | (*)323,100,000 | 333,683,445 | 233,164,871 | 229,668,585 | 186,522,065 | 194,538,374 | 53,187,512 | 39,668,686 | 13,499,502 | 21,922,391 | 9,080,932 | 34 |
| 35 | Internal Revenue — $ | 5,399,189,745 | 289,933,519 | 295,327,927 | 142,606,706 | 124,009,374 | 184,899,756 | . . . . | . . . . | 12,982 | . . . . | . . . . | 35 |
| 36 | Disbursements—ordinary — $ | 6,141,745,740 | 659,705,391 | 487,713,792 | 297,736,487 | 264,847,637 | 293,953,605 | 63,130,598 | 40,948,563 | 24,314,129 | 15,411,598 | 10,830,971 | 36 |
| 37 | War — $ | 1,604,414,792 | 157,911,360 | 134,774,768 | 44,582,838 | 38,116,916 | 57,655,675 | 16,472,202 | 9,687,025 | 7,097,767 | 4,767,129 | 2,460,820 | 37 |
| 38 | Navy — $ | 629.91,116 | 123,173,717 | 55,953,078 | 22,006,206 | 13,536,985 | 21,780,230 | 11,514,650 | 7,904,725 | 6,113,897 | 3,239,429 | 3,448,716 | 38 |
| 39 | Pensions — $ | 213,144,594 | 160,406,416 | 140,877,316 | 106,936,855 | 56,777,174 | 28,340,202 | 1,100,802 | 1,866,886 | 2,603,562 | 1,363,297 | 64,131 | 39 |
| 40 | Interest on public debt — $ | 1,020,281,442 | 21,342,979 | 40,160,333 | 36,099,284 | 95,757,575 | 129,235,498 | 3,144,121 | 3,782,393 | 174,598 | 1,912,415 | 3,402,601 | 40 |
| 41 | Imports of merchandise — $ | 5,278,481,490 | 1,556,947,430 | 849,941,184 | 789,310,409 | 667,954,746 | 435,958,408 | 353,616,119 | 173,509,526 | 98,258,700 | 70,876,920 | 91,252,768 | 41 |
| 42 | Per capita — $ | 47.52 | 18.44 | 19.88 | 12.56 | 13.31 | 11.98 | 11.53 | 7.48 | 5.58 | 4.87 | 17.19 | 42 |
| 43 | Exports of merchandise — $ | 8,108,988,063 | 1,744,984,720 | 1,394,483,082 | 857,878,684 | 835,638,658 | 392,771,768 | 333,576,057 | 144,375,726 | 123,668,932 | 71,670,742 | 70,971,780 | 43 |
| 44 | Per capita — $ | 71.50 | 18.95 | 17.05 | 13.69 | 16.43 | 49.17 | 10.01 | 6.23 | 7.25 | 5.57 | 13.37 | 44 |
| 45 | Imports, silk, raw — Lbs | 30,636,374 | 25,457,273 | 11,256,310 | 5,347,309 | 2,565,716 | 987,504 | . . . . | . . . . | . . . . | . . . . | . . . . | 45 |
| 46 | Rubber, crude — Lbs | 566,346,138 | 101,944,681 | 49,577,108 | 33,442,574 | 16,876,099 | 9,624,998 | . . . . | . . . . | . . . . | . . . . | . . . . | 46 |
| 47 | Tin plates — Lbs | 130,143 | 154,366,999 | 147,961,804 | 680,019,823 | 879,302,882 | 160,912,768 | . . . . | . . . . | . . . . | . . . . | . . . . | 47 |
| 48 | Iron, steel, manufactures of — $ | 50,352,808 | 38,502,457 | 29,478,728 | 41,679,591 | 71,266,099 | 40,773,682 | 20,158,246 | 20,143,087 | 8,157,925 | 6,140,387 | . . . . | 48 |
| 49 | Domestic exports, manufactures — $ | 4,160,554,097 | 766,791,243 | 484,843,299 | 178,932,803 | 121,849,508 | 70,019,445 | 48,475,910 | 23,232,416 | 15,427,140 | 9,879,185 | . . . . | 49 |
| 50 | Iron, steel, manufactures — $ | 1,112,773,886 | 179,143,186 | 121,913,548 | 25,542,538 | 14,710,524 | 15,465,483 | 5,876,154 | 4,960,702 | 1,157,877 | 327,747 | 52,144 | 50 |
| 51 | Farm animals—value — $ | 8,507,148,090 | 5,138,486,000 | 2,228,124,551 | 2,418,766,028 | 1,576,917,550 | 1,584,060,149 | 1,089,329,915 | 544,180,516 | . . . . | . . . . | . . . . | 51 |
| 52 | Cattle — No. | 68,769,000 | 61,803,000 | 47,387,414 | 52,801,907 | 33,258,900 | 23,484,280 | 25,616,019 | 17,778,917 | 14,971,586 | . . . . | . . . . | 52 |
| 53 | Horses — No. | 20,783,000 | 19,833,000 | 13,537,524 | 14,213,837 | 11,201,800 | 8,248,800 | 6,249,174 | 4,446,719 | 4,335,669 | . . . . | . . . . | 53 |
| 54 | Sheep — No. | 47,114,000 | 52,447,000 | 41,883,065 | 44,336,072 | 40,765,900 | 28,477,951 | 22,471,275 | 21,723,220 | 19,311,374 | . . . . | . . . . | 54 |
| 55 | Mules — No. | 5,043,000 | 4,210,000 | 2,086,027 | 2,331,027 | 1,729,500 | 1,126,000 | 1,151,148 | 559,331 | . . . . | . . . . | . . . . | 55 |
| 56 | Swine — No. | 71,757,000 | 58,186,000 | 37,079,056 | 51,602,780 | 34,034,100 | 26,751,400 | 33,512,867 | 30,354,213 | 26,301,293 | . . . . | . . . . | 56 |
| 57 | Production of gold — $ | 49,509,461 | 96,269,100 | 79,171,000 | 32,845,000 | 36,000,000 | 50,000,000 | 46,000,000 | 50,000,000 | 611,687,829 | 604,950 | . . . . | 57 |
| 58 | Silver, commercial value — $ | 57,129,479 | 30,854,561 | 35,741,100 | 57,242,100 | 34,717,000 | 16,434,000 | 156,800 | 50,000 | . . . . | . . . . | . . . . | 58 |
| 59 | Coal — Tons | (*)477,714,821 | 447,849,969 | 240,789,310 | 140,866,931 | 63,822,830 | 29,496,054 | 13,044,680 | 8,266,523 | 1,848,749 | 246,170 | 150 | 59 |
| 60 | Petroleum — Galls | (*)12,865,134,000 | 8,901,404,416 | 2,672,006,218 | 1,924,390,024 | 1,104,037,060 | 220,451,700 | 21,000,000 | . . . . | . . . . | . . . . | . . . . | 60 |
| 61 | Pig Iron — Tons | (*)31,013,564 | 27,303,587 | 13,789,242 | 9,202,703 | 3,835,191 | 1,665,179 | 821,223 | 563,755 | 286,903 | 165,000 | . . . . | 61 |
| 62 | Steel — Tons | (*)134,671,232 | 26,094,919 | 10,188,329 | 4,277,071 | 1,247,335 | 68,750 | . . . . | . . . . | . . . . | . . . . | . . . . | 62 |
| 63 | Tin Plates — Lbs | (*)2,578,411,487 | 1,619,045,000 | 349,901,922 | 2,336,743 | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | 63 |
| 64 | Copper — Lbs | (*)1,509,750 | 182,334 | 270,588 | 115,966 | 57,000 | 12,000 | 7,500 | 850 | 190 | . . . . | . . . . | 64 |
| 65 | Wheat — Lbs | 80,167,000 | 321,062,750 | 288,636,621 | 276,800,000 | 232,500,000 | 162,000,000 | 60,264,913 | 52,516,840 | 37,180,134 | . . . . | . . . . | 65 |
| 66 | Wheat — Bush | 787,184,000 | 635,121,000 | 522,229,505 | 399,262,000 | 498,549,868 | 287,844,199 | 173,104,924 | 100,485,944 | 84,823,272 | . . . . | . . . . | 66 |
| 67 | Corn — Bush | 3,232,367,000 | 2,886,260,000 | 2,105,102,516 | 1,489,970,000 | 1,717,434,543 | 1,094,255,000 | 838,792,740 | 592,071,104 | 377,531,875 | . . . . | . . . . | 67 |
| 68 | Cotton — Running Bales | (*)11,421,352 | 11,568,334 | 10,245,602 | 8,652,597 | 6,605,750 | 4,352,317 | 3,849,469 | 2,454,442 | 1,634,954 | 1,026,993 | 155,509 | 68 |
| 69 | Cane Sugar — Lbs | 348,750,000 | 739,400,000 | 325,749,011 | 281,784,285 | 178,872,000 | 87,044,000 | 230,982,000 | 247,577,000 | 120,471,074 | . . . . | . . . . | 69 |
| 70 | Sugar consumed — Lbs | 9,732,449,275 | 7,460,139,111 | 4,477,175,336 | 3,192,735,908 | 1,870,731,478 | 1,263,941,685 | 767,192,714 | 426,676 | 245,034 | 129,984 | 18,729 | 70 |
| 71 | Cotton consumed — 500 lb. Bales | 6,262,000 | 4,536,729 | 3,680,346 | 2,464,491 | 1,851,922 | 1,096,541 | 845,823 | 635,384,604 | 745,311,061 | 298,189,102 | 17,789,803 | 71 |
| 72 | Domestic cotton exported — lbs | 3,129,134,386 | 3,290,140,326 | 3,109,983,588 | 2,473,700,818 | 1,822,061,114 | 958,554,721 | 1,767,686,338 | 57.92 | 37.68 | 23 | . . . . | 72 |
| 73 | Railways in operation — Miles | 255,264,363 | 240,292 | 198,964 | 166,703 | 93,207 | 52,922 | 30,635 | 9,021 | 2,818 | 23 | . . . . | 73 |
| 74 | Passengers carried — No. | (*)1,084,597,406 | 955,647,199 | 576,831,513 | 492,430,865 | 289,269,603 | 91,675,346 | 4,051 | 2,526 | 1,403 | . . . . | . . . . | 74 |
| 75 | Freight carried one mile — Tons | (*)405,378,741,708 | 255,016,910,451 | 141,596,551,103 | 77,207,672,598 | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | . . . . | 75 |
| 76 | American vessels Built — Tons | 3,880,639 | 342,068 | 393,790 | 294,122 | 157,409 | 276,953 | 214,798 | 272,218 | 118,309 | 58,560 | 106,261 | 76 |
| 77 | Trading, domestic, etc — Tons | 10,375,776 | 6,726,907 | 4,286,516 | 3,409,000 | 2,637,688 | 2,730,797 | 2,546,237 | 1,755,023 | 1,788,999 | 622,475 | 398,977 | 77 |
| 78 | Foreign, and engaged in whaling — Tons | 9,926,390 | 791,825 | 826,694 | 946,695 | 1,314,402 | 1,516,800 | 2,546,237 | 1,585,711 | 899,765 | 576,475 | 669,921 | 78 |
| 79 | Tonnage entered — Tons | 3,158,690 | 7,894,502 | 5,985,322 | 1,005,063 | 685,102 | 484,751 | 467,774 | 188,205 | 54,199 | 11,106 | (*)69,580 | 79 |
| 80 | Through Suez via Marie — Tons | 56,194,083 | 49,650,128 | 32,111,954 | 16,416,435 | 1,774,800 | 4,067 | 64,102 | . . . . | . . . . | . . . . | . . . . | 80 |
| 81 | Commercial failures — No. | 8,881 | 12,652 | 10,774 | 10,907 | 4,735 | 3,546 | . . . . | . . . . | . . . . | . . . . | . . . . | 81 |
| 82 | Amount of liabilities — $ | 295,121,805 | 201,757,097 | 138,495,673 | 189,856,964 | 65,752,000 | 88,242,000 | 79,807,000 | . . . . | . . . . | . . . . | . . . . | 82 |
| 83 | Post office — No. | 52,658 | 59,580 | 75,088 | 62,401 | 42,989 | 28,498 | 18,417 | 13,468 | 8,450 | 8,004 | 903 | 83 |
| 84 | Receipts, P. O. Department — $ | 437,150,312 | 224,128,651 | 102,374,579 | 60,882,098 | 33,315,479 | 19,772,221 | 8,518,067 | 5,499,984 | 4,543,522 | 1,850,583 | 280,804 | 84 |
| 85 | Telephones in use — No. | no data | 7,471,435 | 79,400,322 | 61,754,782 | 29,751,509 | 9,157,646 | . . . . | . . . . | . . . . | . . . . | . . . . | 85 |
| 86 | Newspapers etc. — No. | 24,012 | 22,736 | 20,806 | 16,948 | 9,723 | 5,871 | 4,051 | 2,526 | 1,403 | . . . . | . . . . | 86 |
| 87 | Public school salaries — $ | (*)1,036,477,000 | 253,915,129 | 137,687,746 | 91,836,484 | 55,942,972 | 37,832,566 | 4,778 | 903 | 473 | . . . . | . . . . | 87 |
| 88 | Patents issued — No. | 39,482 | 35,930 | 26,499 | 26,292 | 13,947 | 13,333 | 4,778 | 993 | 473 | . . . . | . . . . | 88 |
| 89 | Immigrants arrived — No. | 430,001 | 1,041,570 | 448,573 | 455,302 | 457,257 | 387,203 | 150,237 | 369,980 | 84,066 | 23,322 | . . . . | 89 |

EXPLANATION OF NOTES IN CONNECTION WITH TABLE OF AMERICAN GROWTH

# INDEX